空间数据多尺度表达的不确定性分析模型

Uncertainty Analysis Models for Multi-scale Representation of Spatial Data

徐丰 牛继强 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

内容提要

本书以建立空间数据多尺度表达不确定性的分析模型为研究目标，以地图学、空间数据与空间分析不确定性原理、图形图像学、模糊数学、粗集理论、拓扑学、计算机视觉、信息熵、统计学和比较学等多个学科领域知识为基础，紧密结合生产实际，对空间数据多尺度表达的不确定性及其评价指标、量化模型，空间数据多尺度表达的空间关系不确定性及其评价指标、量化模型，空间数据多尺度表达不确定性可视化方法等进行了系统深入的研究。全书共分为7章，主要内容包括：空间数据多尺度表达不确定性相关概念及相关研究进展，空间数据多尺度表达不确定性的理论基础，空间数据多尺度表达不确定性及其形式化描述，空间数据多尺度表达的不确定性建模，空间数据多尺度表达的空间关系不确定性建模，空间数据多尺度表达的不确定性可视化，结语。

本书可作为地球空间信息科学、地理学及相关专业的研究人员、高校教师、研究生的参考用书，也可供相关工程技术人员参考使用。

■ 责任编辑／鲍
■ 责任校对／鄢
■ 版式设计／韩
■ 封面设计／汪

ISBN 978-7-307-12800-2

定价: 39.00元

国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研
国家自然科学基金项目(41201387)
信阳师范学院第六批一级重点学科——地理学
河南省高等学校青年骨干教师资助计划(2012-GGJS-127)
联合资助

空间数据多尺度表达的不确定性分析模型

Uncertainty Analysis Models for Multi-scale Representation of Spatial Data

徐丰 牛继强 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

空间数据多尺度表达的不确定性分析模型/徐丰,牛继强著.—武汉:武汉大学出版社,2014.4

ISBN 978-7-307-12800-2

Ⅰ.空… Ⅱ.①徐… ②牛… Ⅲ.空间信息系统—不确定系统—系统模型—研究 Ⅳ.P208

中国版本图书馆 CIP 数据核字(2014)第 020475 号

责任编辑:鲍 玲　　责任校对:鄢春梅　　版式设计:韩闻锦

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:武汉中远印务有限公司

开本:720×1000 1/16 印张:17.25 字数:243 千字 插页:1

印数:1－1050 册

版次:2014 年 4 月第 1 版 2014 年 4 月第 1 次印刷

ISBN 978-7-307-12800-2 定价:39.00 元

前　　言

随着 GIS 应用领域的不断扩展和需求层次的日益提高，人们越来越多地需要对多尺度的空间数据进行分析、处理和表达。如今，空间数据的多尺度表达已成为地理信息科学研究的前沿问题之一。不确定性问题存在于空间数据的生命周期全过程，只要有空间数据就有不确定性问题存在。因此，空间数据的多尺度表达必然存在不确定性。

空间数据的不确定性问题早已引起国内外 GIS 界的重视，是当前 GIS 理论研究的热点和难点之一。国外对于空间数据的不确定性问题非常重视。美国地理信息科学大学研究中心（UCGIS-CI）、美国国家地理信息和分析中心（NCGIA）、美国国家经济与社会研究委员会（ESRS）、荷兰的空间数据分析专家中心（NEXPRI）都把地理信息的不确定性问题作为重要课题进行研究。在国际空间数据处理会议（ISDH）、欧洲地理信息系统会议（EGIS）、美国地理信息系统年会（AGIS）、自然资源数据库空间数据不确定性等国际会议中，都设立了关于空间数据不确定性专题讨论组。以空间数据不确定性为主题的国际空间数据质量大会已经召开了 8 届；两年一届的自然资源与环境科学中的空间精度评价国际会议已经召开了 10 届。国内外对空间数据不确定性的研究取得了长足进步，主要从"GIS 不确定性的研究内容与体系、数据不确定性的来源、数据不确定性处理方法、空间分析过程中的不确定性传播机理"等方面展开。目前国内已有较为系统的研究和出版的专著，如史文中（2005）的《空间数据与空间分析不确定性原理》，承继成、郭华东、史文中等人（2004）的《遥感数据的不确定性问题》，邬伦、高振记等人（2010）的《地理信息系统中的不确定性问题》、胡圣武（2006）的《GIS 质量

评价与可靠性分析》等。

关于空间数据多尺度表达不确定性的问题，早已引起学者的关注，Lam 等人在 1992 年将尺度和数据不确定性之间的关系归纳为四个方面：不同的量测单位其结果不一样；不同比例尺具有不同的不确定性；大尺度的研究覆盖较大的研究区域，代表性强，不确定性小；运行尺度越小把握性越大，不确定性越小。随着空间数据多尺度表达问题的升温，空间数据多尺度表达的不确定性问题也越来越引起学者们的重视，第八届(2013)国际空间数据质量大会上有数篇空间数据多尺度表达不确定性相关的论文参与交流，在自动综合算法的质量评价中也都包含了对综合结果不确定性的评估，而且许多空间数据不确定性的研究与空间数据多尺度表达不确定性的研究相关，其成果可为空间数据多尺度表达不确定性研究利用和借鉴。

在国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研究”(No. 41001219)的资助下，本书作者对空间数据多尺度表达的不确定性及其评价指标、量化模型，空间数据多尺度表达空间关系不确定性及其评价指标、量化模型，空间数据多尺度表达不确定性可视化方法等进行了系统深入的研究，取得了一些阶段性成果。本书除了对这些成果进行全面的总结，还借鉴了国内外相关领域的最新研究进展，力图对空间数据多尺度表达不确定性的理论和方法进行系统的阐述，初步建立空间数据多尺度表达不确定性研究的理论和方法体系。

本书共分为 7 章。第 1 章(绪论)阐明空间数据多尺度表达不确定性的相关概念及相关研究进展；第 2 章(空间数据多尺度表达不确定性的理论基础)介绍相关数学基础以及空间数据不确定性模型；第 3 章(空间数据多尺度表达不确定性及其形式化描述)研究空间数据多尺度表达不确定性的来源和分类，并归纳不同尺度空间数据不确定性的统一的形式化描述方法，作为开展相关研究的基础；第 4 章(空间数据多尺度表达的不确定性建模)深入研究各地理要素多尺度表达的不确定性评价指标和量化模型；第 5 章(空间数据多尺度表达的空间关系不确定性建模)深入研究拓扑关系和方

向关系多尺度表达的不确定性评价指标和量化模型；第6章(空间数据多尺度表达的不确定性可视化)系统研究空间数据多尺度表达不确定性的可视化方法；第7章(结语)对本书进行总结。

本书由信阳师范学院城市与环境科学学院的徐丰、牛继强共同拟定提纲；徐丰撰写了本书的前言、第1章、第2章的第1节、第3章、第4章的第1节和第2节、第5章的第1节和第2节等几个部分，牛继强撰写了本书的第2章的第2节和第3节、第4章的第3节至第6节、第5章的第3节至第5节、第6章、第7章等几个部分。李卓凡、蔡珺同学在整理资料、绘制图表和校对方面做了大量的工作，王帅帅、王腾、付素华等同学也参与了资料整理与校对工作。最终，由徐丰、牛继强统稿。在本书的研究和撰写过程中，得到了信阳师范学院城市与环境科学学院领导的关怀和同事的支持，在此表示诚挚的谢意。

本专著是在国家自然科学基金项目“空间数据多尺度表达不确定性及其可视化研究”(No. 41001219)的资助下取得的研究成果之一，并得到国家自然科学基金(No. 41201387)、信阳师范学院第六批一级重点学科——地理学一级重点学科和河南省高等学校青年骨干教师资助计划(No. 2012-GGJS-127)的联合资助。

由于研究问题的复杂性和研究时间的限制，本专著的研究仍是初步的，在一些方面尚不成熟，错误和疏漏在所难免，敬请读者不吝指正。

徐 丰 牛继强

2013年12月于信阳师范学院

目　录

第1章　绪　论

空间数据的多尺度表达与GIS数据的精度分析和质量控制问题是当今地理信息科学和国际GIS界十分关注的两大前沿课题。数据质量的关键是对数据不确定性的评估。许多著名的国际学术机构和组织都把空间数据的不确定性问题作为21世纪的重大研究课题。目前，GIS不确定性的研究集中在GIS基本实体(点、线、面等)的单一数据集的不确定性的研究上，很少涉及多尺度表达(数据转换)的空间数据的不确定性(可靠性)评价的模型、度量指标和评价体系。因此，为了适应地理信息数据多尺度表达的发展需要，必须研究空间数据多尺度表达不确定性评价模型、度量指标和评价体系，一方面为正确、合理、有效地建立GIS多尺度表达实用系统提供依据和标准，另一方面为客观、科学地理解、分析和使用多尺度表达的空间数据提供基本保障，降低空间决策的风险。

1.1　基本概念

1.1.1　空间数据

“空间”的概念在不同的学科有着不同的解释(王家耀，2001)。地理学中的地理空间是指物质、能量、信息的存在形式在形态、结构过程、功能关系上的分布方式和格局及其在时间上的延续。地理空间涵盖上至大气电离层、下至地幔莫霍面的范围，是地球上大气圈、水圈、生物圈、岩石圈和土壤圈交互作用的区域。依附地理空间存在的各种事物或现象，它们可能是物质的，也可能是非物质

的，如土地类型、资源分布、行政区划、水系、人口分布、工农业布局、城市规划、道路网结构等。这些事物和现象的一个典型特征，就是与一定的地理空间位置相关，都具有一定的几何形态，因此称为地理空间实体。在地理空间中，实体不仅反映了事物和现象的地理本质内涵，而且反映了它们在地理空间中的位置、分布状况以及它们之间的相互关系。

地理信息科学中的地理空间被定义为绝对空间和相对空间两种形式，我们称之为地理信息空间。绝对空间是具有属性描述的空间位置的集合，它是由一系列不同位置的空间坐标值组成；相对空间是具有空间属性特征的实体的集合，它是由不同实体之间的空间关系构成。可以简单地将地理信息空间理解为一个空间目标(对象)组合排列集，其每个目标都具有位置、属性和时间信息，及与其他对象的拓扑关系、语义关系等。

从地理信息空间的定义可知，它是信息世界的地理空间，是地理学地理空间的抽象表达，可以通过对地理空间进行抽象或空间采样来描述，其结果就是地理空间数据。因此，地理空间数据是地理空间信息的数字描述，是地理空间的离散表达，由空间对象的几何信息、属性信息、时间信息，以及与其他对象之间所具有的拓扑关系和语义关系等数据构成。我们可以把这些数据分为三类：属性数据、空间数据以及时域数据，其中空间数据是指含有空间或位置信息的数据，分为关系数据和几何数据两种。

关系数据是描述各个不同空间实体之间关系(如邻接、关联、包含、连通、接近度)的信息，也就是拓扑关系、方向关系、距离关系。几何数据从几何角度把空间目标划分为点状、线状和面状目标三种基本类型。点状(零维)目标，是在空间有确定的位置(坐标)而没有长度和面积的目标。例如，孤立点、拓扑交点和端点。线状(一维)目标，是在空间有确定位置和长度而没有面积的目标，其端点由两个点状目标界定。面状(二维)目标，是在空间有确定的位置，并有长度和面积的目标，它由若干个点状或线状目标界定。图1.1为空间数据的分类。

无限的地理实体离散化会产生有限的表示，因为计算机模型和

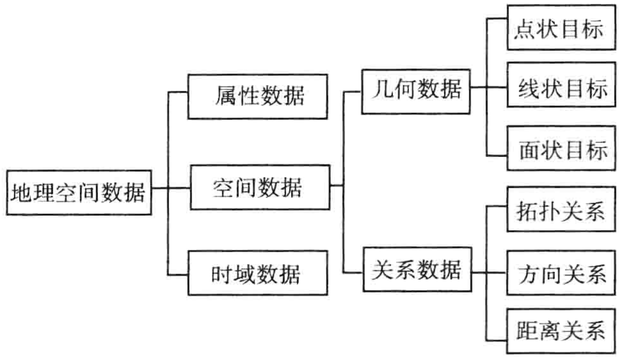


图 1.1 空间数据的分类

数据获取方法是有限的，这个离散化是通过两个步骤来完成的：分辨率控制和特征抽象。按照所表达地理实体的知识，利用分辨率可控制其详细程度(Bruegger，1994)。依据其详细程度，空间数据有三个水平上层次结构：要素层、目标层、几何细节层。要素是具有相同语义特征的目标集；目标是具有独立地理意义的表达实体，是构成要素的基本单位；几何细节是几何表达上划分的结构体，是构成目标的基本单位，如构成河流目标的“弯曲”特征，构成面状目标的三角形剖分单元等。

1.1.2 空间数据多尺度表达

尺度(scale)是易混淆和易误解的概念，通常对于不同的学科和应用背景有不同的含义(Goodchild & Quattrochi，1997)。尺度是指在研究某一物体或现象时所采用的空间或时间单位，同时又可指某一现象或过程在空间和时间上所涉及的范围和发生的频率。前者是从研究者的角度来定义尺度，而后者是根据所研究的过程或现象的特征来定义尺度。

尺度也是与地理信息相关的最基本的和难以理解的概念之一。在地理信息科学中，尺度既用来指研究范围的大小(如地理范围)，也用于指详细程度(如地理分辨率的层次、大小)，还用于表明时

间的长短以及频率(即时间尺度)。

在研究的范围大小上(即空间尺度)，往往以粒度(grain)和幅度(extent)来表达。空间粒度是指地理空间中最小可辨识单元所代表的特征长度、面积或体积(如样方、像元)。对于空间数据如影像资料而言，其粒度对应于像元大小，与分辨率有直接关系。幅度是指研究对象在空间或时间上的持续范围或长度。具体地说，所研究区域的总面积决定该研究的空间幅度。

在时间尺度上，时间粒度是指某一现象或事件发生的(或取样的)频率或时间间隔。例如，一些测量仪器的取样时间间隔或某一干扰时间发生的频率。一般根据研究项目持续的时间长短来确定其时间幅度。

关于详细程度，可以用人们所熟悉的地图比例尺概念来理解。“比例尺”定义为表达空间(地图图面)中的距离与实际地理空间距离的比率。大比例尺的地图覆盖了较小的地理区域并一般包含了较为详细的信息；相反，小比例尺的地图覆盖了较大的地理区域，但通常所包含的有关该区域的信息较少。不同的比例尺所表达的信息密度有很大的差异，通常以不同分辨率或抽象层次来表达同一地理要素不同详细程度的几何、拓扑结构和属性方面的信息。王艳慧、陈军等人(2003)对此问题作了非常形象直观的解释，在几何层面上，同一地理要素在不同的尺度下表现出不同抽象程度的几何形状，反映在数据库中则可能被抽象为不同的几何类型。如图1.2所示，道路网中同样几条路段，在较大比例尺下，各路段表达为双线目标，道路节点用面状几何要素表示；但在小比例尺下，同样的路段则用其道路中心线表示，立交桥、路口等节点则用简单的点状目标来表示。在要素层面上，几个要素可在不同的抽象层次下，基于不同的几何、时态或语义准则聚合成新的复合要素。这样的一个复合要素可能在不同尺度下存在不同的表达，且各种表达相互独立，在不同比例尺转换时会发生出现/消失或聚集/分解的情况。这种情况下，复合要素和底层的对应要素间具有层次性关系，高层要素由低层要素组合而成。如图1.3所示，河南省行政区划要素由17个市(地区)级行政区划要素组成，后者又由若干个县(市、区)级行

政区划要素组成，彼此之间存在行政上的隶属关系和空间上的聚集/分解关系。在属性层面上，同一地理要素在不同抽象层次下的表达属性也不相同。表1.1中，点、线、面在不同尺度背景(村级、乡(镇)级、县(市、区)级、市(地区)级、省级、国家级)下反映的要素的属性具有不同的含义。低层抽象对象的属性值比高层对象的属性值更准确。与前面对象层次性不同的是，要素层次上各尺度下的抽象对象是客观存在的，而属性值层次中的对象实际上并不存在，且属性值与尺度(分辨率)有关。

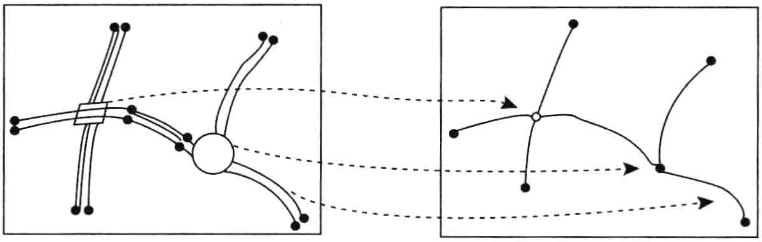


图1.2 地理要素多尺度集合抽象表达的层次性



图1.3 地理要素不同尺度抽象层次下的要素聚集特征

在地理信息系统的研究中，一般将尺度简单地理解为比例尺。不同的尺度不仅所表达的信息在密度上有很大的差异，而且还会影响所表达的地理信息是否正确，因为不少地学现象和规律只在一定的尺度出现，只有经过合理的尺度抽象出的空间信息才更具利用价值。因此，从认知科学的观点出发，尺度体现了人们对空间事物、空间现象认知的深度与广度。

表 1.1　同一几何抽象要素在不同尺度下反映不同详细层次的属性信息

要素	点	线	面
村级	位置、点属性	位置、形状、方向	位置、形状、面积
乡(镇)级	位置、点属性	位置、形状、方向	位置、形状、结构、面积
县(市、区)级	位置、点属性	位置、形状、方向	位置、形状、结构、面积
市(地区)级	位置、点属性	位置、走势	位置、形状、结构
省级	位置、高级别点属性	轮廓、走势	位置、结构
国家级	位置、高级别点属性		轮廓

李霖等人(2005)将尺度定义为与空间数据的空间、属性、时间特征同等重要的基本特征。空间数据的多尺度表达就是要解决由于用户对地理数据不同应用的需求，而导致需要对空间数据产生不同表达的问题。王艳慧等人(2006)认为地理要素多尺度表达指的是在数据库中存储、定义、描述、维护地理要素两个或两个以上尺度(比例尺)下的相应表达实例，这些实例以不同分辨率或抽象层次来表达同一要素不同详细程度的几何、拓扑结构和属性方面的信息，使得用户可以根据具体需求选择其中的一种或多种不同分辨率的实例，并能用于多比例尺的层次空间推理和数据更新。由此可见，空间数据的多尺度表达是在不同尺度(比例尺)下以相适应的不同详细程度来表达空间数据，一个特定尺度下的空间数据可以转换到另一个尺度下的空间数据，这种转换可以是向上尺度转换(尺度扩展)，也可以是向下尺度转换(尺度收缩)。常用关于尺度转换

(孟斌, 王劲峰, 2005)的表述见表 1.2。

表 1.2 常用关于空间尺度转换的表达

	英文表达	其他中文翻译	具体含义
尺度扩展	Upscaling Scaling-up Top-down	向上尺度转换 尺度上推	从小尺度(空间)观测中获得较大尺度(空间)上信息的过程，是将信息从精确的尺度(高分辨率)向模糊尺度转换的过程
尺度收缩	Downscaling Scaling-down Bottom-up	向下尺度转换 尺度下推	将大尺度上(空间)的信息分解到更小尺度(空间)上的过程，是将信息从模糊的尺度向精确的尺度转换的过程

1.1.3 不确定性

"不确定性"是相对于"确定性"而提出来的，"不确定性"是客观世界的固有属性，它遍及万事万物和一切客观现象中。

比较常见的关于不确定性的定义有：不确定性(uncertainty)是指客观世界或实体本身就具有的变异，表现为不精确性、随机性和模糊性；不确定性是指既不能肯定，也不能否定，或可能是对或可能是错，没有准确把握；不确定性表示事物的含糊性、不明确性、不肯定性或指某事物的未决定或不稳定状态；混沌(chaos)理论认为，复杂系统中的不确定性是由于系统中各要素的相互作用一般处于力的不平衡状态，在空间上表现为要素的不均匀性和不稳定性，在时间上则表现为发展的非周期性，它们综合反映在客观实体上就是对象的不确定性(spatial object with uncertainty)。总括前人的研究成果，客观世界的不确定性问题可归为客观世界客观存在的不确定性和人们主观认识上的不确定性两类。客观世界客观存在的不确定性最为典型的是随机性；人们主观认识上的不确定性有：由于定义不清造成的模糊性、灰色系统中的部分已知、部分未知的灰性，以

及未确知性。其中，未确知性是指就客观世界本身来讲它是确定的，但对于决策者来说由于受决策条件限制而导致的认识不清、所掌握的证据不足以确定事物的真实状态和数量关系而带来的纯主观的、认识上的不确定性。

不确定性的含义比较广泛，以下现象均具有不确定性：

①凡是既不能肯定，也不能否定的，为不确定性。

②凡是一个图形、影像、数据或方程具有多解性的，该方程或数据具有不确定性，如遥感数据、地球物理数据等。

③在一个多尺度系统中，物理量是随尺度而变化的。例如，英国或挪威的海岸线不仅随度量的尺度的不同而不同，而且还随比例尺的不同而不同，在1∶50万和1∶5万地图上测量的长度，除了规定的地图比例差之外，还存在“多尺度系统的不确定性问题”。

④凡是瞬息万变的动态系统，其变化的速度大于仪器可能测量的速度，它是测不准的，是不确定的。

⑤人工模拟产品与客观真实世界之间总存在一定的差异，没有差异是不可能的，也就是具有不确定性。

⑥凡是两类地物之间是渐变的，则两者之间的界线具有不确定性。

⑦凡是概念或语义具有不完备性或不一致性特征的，则具有不确定性。例如，城市的定义在各国是不一致的，所以城市化率的概念也是不确定的；林地在农业部、林业部、土地局的定义可能是不一样的，所以是不确定的。

数据是对客观世界性质、特征和状态的描述，但是由于客观世界的复杂性和在数据产生过程中携带了一些和客观世界无关的干扰影响，使得数据产生了与客观世界不一致的状况，这种不一致导致了数据具有不确定性。数据的不确定性大概由三方面的因素所致：①客观世界本身的不确定性；②人类对客观世界认知的局限性；③量测技术水平及手段的局限性。

客观世界中的绝大部分现象都是不确定的。所谓确定的、规则的现象，只会在一定的前提和条件下发生，只会在局部或者较短的时间内存在。例如，在构成物质的分子内部，其电子围绕原子的运

动轨迹是不确定的；城市与郊区的结合部都明显地呈现边界的不确定性；气候分布、各种林地分布的边界也都呈现出不确定性特征。

人类对客观世界的认知是一个渐进的逐步深入的漫长过程，而这种认知是渐进的、不断深入发展的，在某一时期总是处于一定的水平，并且在认知进程的每一阶段，都无可避免地存在着认知的不确定性。例如，随着航天科技与遥测技术的飞速发展，人们对宇宙的认知进一步深化，其认知的广度是以往任何时期都无与伦比的。

我们对客观目标使用某一量测技术在一定的时期内可能只能达到一定的准确度，而且不同的量测技术可能会导致不同的准确度。例如，使用全球定位系统对地面点坐标进行量测，在具有 SA (selective availability) 政策限制下，定位精度为 100m，取消 SA 限制后定位精度可以达到 20m (实际可能达到约 10m)；而使用差分 GPS 技术，在有足够多的卫星观测值的情况下可以达到 1 ~ 5m 的精度。再进一步，GPS 载波相位定位可达到 1 ~ 2cm，甚至毫米级精度。

“不确定性”存在于自然科学技术的各个领域，如微电子分子生物学、基因工程及医学中，同时还存在于社会经济和人文科学的各个领域。总之，客观世界和我们的生活里充满了不确定性。随着不确定性研究的深入，世界的不确定性特征越来越得到学术界的普遍认可。越来越多的科学家相信，不确定性是这个世界的魅力所在，只有不确定性本身才是确定的。

1.2 空间数据多尺度表达的不确定性问题

1.2.1 空间数据的不确定性

数据是客观实体信息的载体，是科学的基础。空间数据是 GIS 的基础数据。

空间数据通过描述地理空间实体及其实体间各关系来表达和模拟客观地理世界。但是由于主客观的因素，如测量设备的固有精度

以及测量误差，或样点选择、采样间隔等，即使是1∶1的采样和表达，也是对地理空间的综合或抽象，是对客观地理世界的一种近似。

空间数据及其不确定性是两位一体的问题。不确定性必须依附于数据而存在，而数据的不确定性是其主要特征之一。空间数据的不确定性从数据采集过程就开始并贯穿于数据的整个流程中，有的差错可能被发现并得到校正，但新的差错又可能产生，使得不确定性问题存在于整个数据的生命周期(data life cycle；Goodchild，2000)。因此，不确定性问题存在于空间数据的生命周期全过程，在其生命期的每一个环节(地理现象认知、空间数据获取、空间数据预处理和空间分析等)都可能存在着不确定性。图1.4为一般空间数据流程框架，同时展现了空间数据不确定性在空间数据流程中的流程。通过分析图1.4，空间数据的不确定性来源可分为以下三类：

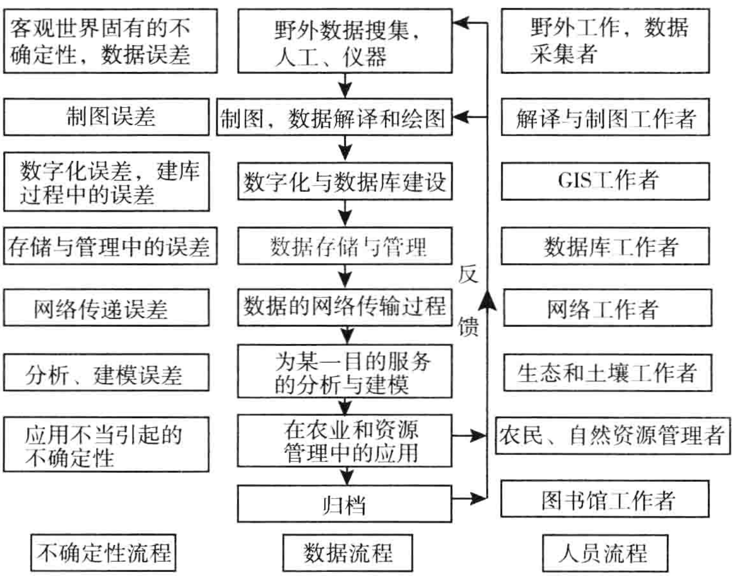


图1.4　一般空间数据流程图

①客观世界固有的不确定性。客观世界的复杂性是造成不确定性的重要原因之一。各领域都有自己固有的不确定性。例如，对真实世界的量测来说，测得准与测不准并存是一个特征。任何仪器都存在不同程度的误差。问题的核心是"真值"是什么。

②人为引起的不确定性。人为引起的不确定性包括认知过程、科学技术的应用、模型、公式计算及操作过程等引起的不确定性。其中，以认知过程最为重要，它是科学技术应用、建立模型和公式的基础。人们认知过程的复杂性与客观世界的复杂性两者的耦合就决定了人为引起的不确定性，其中以认知过程的复杂性为主。在不确定性的总量中，人为引起的不确定性占70%，因此，它是最主要的，从而也引起了人们对它的广泛关注。

③地理空间数据不确定性的其他问题。如地理数据大多为近似值，很少有真值；地理数据很多是衍生数据等。

Gottsegen 和 Goodchild(2001)也指出，空间信息的不确定性主要体现以下三个方面：

①不确定性在一定程度上代表了获得的数据与真实世界之间有明显的质量差别。这些差别可能是由于以下原因引起的：

a. 量测误差(偏差、随机误差)；

b. 没有充足的知识去量测具有精确定义的概念；

c. 概念不能被清楚定义；

d. 量测准确度的局限性或存贮设备精确度的局限性。

②不确定性与用户的专门需求和数据之间的相容性及协调性有关，包括：

a. 词汇(术语)上的不同(命名上的不同)；

b. 语义上的不同(同一术语的不同解释)；

c. 在分类方面的不同；

d. 数据的几何和逻辑特征的不同(由于比例尺和分辨率的原因)。

③不确定性和用户专门要求与数据之间的匹配有关。

例如，用户要求进行生态建模，但这是一个十分复杂的问题，很难弄清楚需要什么样的数据与专门要求(生态建模)的数据之间

的匹配及数据与数据之间的关系，于是就形成了与建模有关的风险或不确定性。

空间数据不确定性的形成原因是多种多样的，而最终空间数据表现出来的不确定性可以看作是一种广义的误差，它包括可度量和不可度量的误差，也包括数值和概念上的误差。空间数据的随机性、数据和概念的复杂性、模糊性及不完整性，可以视作不确定性问题的内容。可见，空间数据不确定性研究的内容十分广泛。但一般都认为，研究不确定性的度量指标、表达形式和传播规律是空间数据不确定性理论研究的核心问题。邬伦(2010)按照研究内容将空间数据的不确定性分为位置不确定性、属性不确定性、时间不确定性、逻辑关系不确定性、模型不确定性等，按照研究内涵可将空间数据的不确定性分为四个方面，即误差、模糊、歧义和不一致。空间数据不确定性的表达在其表达对象和内容上，主要有对空间实体点、线、面、实体空间关系以及遥感影像的表达。

目前，GIS中的数据不确定性主要包括以下五个方面(承继成等，2004；史文中，2005)：

①位置不确定性，指空间数据库中的信息对实际空间位置的正确表达程度。

②属性不确定性，指空间数据库中的信息对实际空间属性的正确表达程度。

③完整性，指空间数据集是否是期望表达实体的完整表达。

④逻辑一致性，指空间数据库是否自协调一致。

⑤现势性，指空间数据的现势性。一般是通过数据的采集时间、更新时间和更新频度来表现的。

1.2.2　空间数据多尺度表达的不确定性

分形理论或不规则几何理论认为，在一个多尺度的系统中，物理量是随尺度变化而变化的，“一种尺度，一个世界”。即使对于同一对象来说，它也随空间、时间和光谱(属性)尺度的变化而变化，所以是不确定性的。海岸线的长度是随其测量尺度的变化而变化的；气候变暖或变冷是随百年、千年或万年的时间段而测定的；

对于同一对象来说，运用可见光波段或用红外波段所观察到的结果是不同的，所以也是不确定的。Lam 等人在 1992 年将尺度和数据不确定性之间的关系归纳为四个方面：不同的量测单位其结果不一样；不同比例尺具有不同的不确定性；大尺度的研究覆盖较大的研究区域，代表性强，不确定性小；运行尺度越小把握性越大，不确定性越小。

在 GIS 环境下，尺度的含义一般是指信息被观察、表示、分析和传输的详细程度。在不同尺度背景下，地球空间现象或实体往往具有不同的空间形态、结构和细节，为了满足不同领域不同层次用户空间分析的需要，人们需要对地球表层各种自然和人文现象的空间形态结构进行不同程度的抽象，从而形成对空间数据不同尺度下的抽象表达。艾廷华等人(2010)根据空间数据在尺度上展示出的不同形式的演变，以尺度为主线，建立了尺度空间中 GIS 数据表达的生命期模型，图 1.5 为河流表达的生命期示例。在尺度空间 GIS 数据表达的生命期模型中，空间数据的表达具有多态性，在某个尺度点的表达状态过渡到相邻的下一个尺度点的表达状态发生了连续的缓慢变化或者突变。该模型形象直观地展现了不同尺度空间数据的差异，这种差异正是空间数据多尺度表达的不确定性问题。

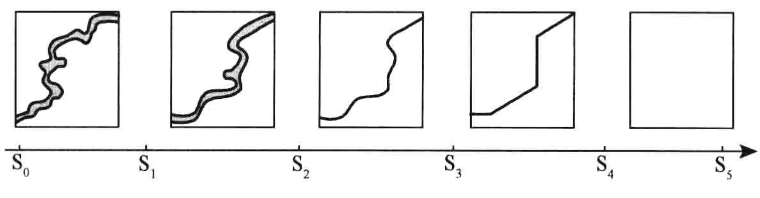


图 1.5　河流表达的生命期示例

不确定性是空间过程和特性不能被准确确定的程度，是自然界各种空间现象本身固有的属性。至今，不确定性还没有一个统一的定义，其含义比较广泛，数据的误差、数据和概念的模糊性及不完整性都可视作不确定性问题研究的内容。因此，空间数据多尺度表达的不确定性，其含义也非常广泛，要给出其明确的定义，是非常

困难的。但是，因为本书所定义的空间数据的多尺度表达为空间数据多比例尺的表达，所以，这里将空间数据多尺度表达的不确定性定义为多种比例尺表达的空间数据与空间现象本身固有的属性之间的差异，重点研究初级尺度空间数据的不确定性对多级比例尺表达的传播、多级比例尺的空间数据与初级尺度的空间数据之间的差异（自动综合引入的不确定性）、空间数据的不确定性对空间关系的传播、多级尺度空间关系的不确定性等。

1.3　空间数据多尺度表达不确定性的研究意义

空间数据的多尺度表达是GIS实用化的核心问题之一，而不确定性制约着GIS使用的可靠性。因此，对空间数据多尺度表达的不确定性的研究有非常重要的理论和现实意义。

①对空间数据多尺度表达不确定性的研究是完善空间数据多尺度表达及空间数据不确定性基础理论的需要。

空间数据的多尺度表达及GIS数据的精度分析和质量控制问题是当今地理信息科学和国际GIS界十分关注的两大前沿和基础性课题。相关机构和学者在这两方面展开了大量的研究，而有关空间数据多尺度表达的不确定性问题的研究相对贫乏。因此，对空间数据多尺度表达不确定性的研究是完善空间数据多尺度表达及空间数据不确定性基础理论的重要内容。

②对空间数据多尺度表达不确定性的研究是认知空间数据多尺度表达不确定性的需要。

不确定性存在于空间数据多尺度表达的整个生命周期，只有对空间数据多尺度表达的不确定性进行系统的定性分析、科学的量化计算，并从视觉的角度表现不确定性信息，才能使GIS用户对其不确定性有充分的认识，并帮助GIS用户了解和解决与不确定性有关的问题。

③对空间数据多尺度表达不确定性的研究是对空间数据多尺度表达产品进行质量评价和质量控制的需要。

空间数据多尺度表达成果是GIS产品的一种重要形式。当前，

GIS 软件设计的假设前提是空间数据中不含有误差，并且 GIS 主要处理确定性数据。这显然与空间数据不确定性的普遍存在性相抵触。Alber(1987)十分尖锐地指出：由于现有的 GIS 不能处理数据及模型和空间操作中的不确定性问题，虽然它能以相当快的速度生产各种表面上看来精美无比的产品，但实际上是一堆废物。因此，GIS 的发展，必须要研究空间数据多尺度表达的不确定性理论，并建立和发展空间数据多尺度表达不确定性的评估方法，为空间数据多尺度表达产品的质量评价和质量控制提供方法和依据。

④对空间数据多尺度表达不确定性的研究是降低或消除决策分析中不确定性风险的需要。

空间数据多尺度表达的主要目的是进行空间数据分析，为空间决策提供信息。在利用空间数据辅助人类决策的过程中，不确定性是广泛存在的。分析、控制和管理不确定性的最终目的，就是帮助数据使用者理解不确定性的存在，并将不确定性所带来的风险减至最低。在决策分析时，如果把数据中的不确定性减少和控制到一个可以接受的水平，那么相应的决策质量就可以得到提高。因此，空间数据及其多尺度表达的不确定性可以直接或间接地影响最终决策的准确性和可靠性。

1.4 相关研究进展

1.4.1 不确定性的认知历程

“不确定性”是相对于“确定性”而提出的。“不确定性”是客观世界的固有属性，它遍及万事万物和一切客观现象中。

在量子力学中，德国物理学家海森堡(Werner Heisenberg)于1927 年提出了“测不准原理”、“非定域性理论”和“电磁波波粒二象性”的“波动数几率及光子的不确定性原理”等。他指出，当人们在测量一个非常小的物体的速度时，其结果总会出现不确定性，即Heisenberg“不确定性原理”。

随后，在数学领域，Gödel（1931）提出了"不完备定理"（incompleteness theorem），这是"关于数字确定性丧失的论证"（承继成等，2003）。Gödel 指出，在任何公理化形式维中，总存在着在定义该维的公理基础上既不能肯定、也不能否定的问题，即"不确定性"问题。"不完备定理"的提出对以后的数学、计算机科学和哲学的发展都具有深远的影响。

在热力学领域，第二热力学——耗散结构创始人、比利时科学家 I. Prigogine 阐述了不确定性的普遍性，指出了经典物理学中的确定性都是有条件的。Prigogine 认为，我们生活在一个不确定的概率世界中，生命和物质在这个世界里沿时间方向不断演化，确定性本身才是错觉；并进一步指出，从宇宙学到分子生物学的所有存在的层面上，都产生了不确定性的演化模式。Prigogine 指出，在时空范围内不存在稳定的基态，没有平衡条件。因而宇宙没有开始，也没有终结，并认为科学只能提供可能性而不能提供确定性。爱因斯坦说时间是一种错觉，Prigogine 认为确定性才是错觉。

在社会科学领域，1945 年以后社会科学家们从坚持决定论的社会体系这一假设很快过渡到承认"社会在本质上是一个不确定的领域"这一常识性命题，普里高津（Prigogine）和许多其他人认为这一常识性命题是一种科学命题，是我们社会科学研究的基础。在诸多的规章制度制定机构、行政部门、国家实验室、大学、企业、国会中，有识之士正在得出一个相同的结论："……忽视这些不确定性的确是不安全的；它对于我们必须做出的决策或许是非常重要的。"

德国物理学家 Haken 于 20 世纪 70 年代初期创立"混沌论"与"协同论"时就指出：按照热力学基本规律，宇宙的混沌特征应该在不断增加，一切有规律的功能顺序应该会停止，一切有序性应该会崩解。同时他还指出，客观世界中存在着有序的经常变换现象，有序的经常变换等于无序。

经典力学是对客观世界的确定性描述，认为客观世界是有序的、有规律的。但是，这种有序性是基于某种条件的确定性。众所周知，1972 年发射的"先驱者 10 号"航天器，其运行轨迹出现怪

异，这使得美国宇航局（NASA）的科学家认为牛顿的“万有引力定律”存在局限性——在宇宙尺度水平上，牛顿的“万有引力定律”不再有效。

爱因斯坦的广义相对论在解释宏观现象时很有效——能很好地解释恒星的运行规律，但无法解释暗物质、暗能量和宇宙“大爆炸”等现象。而在这一方面，Heisenberg 的“不确定性原理”却为其打开了方便之门。

1.4.2 地理空间数据不确定性研究进展

国外对于空间数据的不确定性问题非常重视。1996 年，美国地理信息科学大学研究中心（UCGIS-CI）发表了《地理信息科学的优先研究领域》，该文将“地理数据和基于地理信息系统分析中的不确定性”问题作为重中之重的研究课题；1998 年，美国国家地理信息和分析中心（NCGIA）在制订的研究计划中把空间数据精度和不确定性列为首要内容；美国国家经济与社会研究委员会（ESRS）在制订全面的研究计划时将“地理信息系统中的误差传播”作为重点课题之一；荷兰的空间数据分析专家中心（NEXPRI）在制订的研究计划中也提出了“空间分析理论——空间操作中的误差传播”问题。在国际空间数据处理会议（ISDH）、欧洲地理信息系统会议（EGIS）、美国地理信息系统年会（AGIS）、自然资源数据库空间数据不确定性等国际会议中，都设立了关于空间数据不确定性专题讨论组。

以空间数据不确定性为主题的国际空间数据质量大会（The International Symposium on Spatial Data Quality）已经召开了 8 届；两年一届的自然资源与环境科学中的空间精度评价国际会议（The International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Science）已经召开了 10 届。这两个国际会议的基本内容是一致的，但组织者不同。“Accuracy 2000”国际会议的主要内容包括：

- 误差-灵敏度 GIS 的发展；

- 空间建模的不确定性传播理论和应用；
- 遥感影像分类的不确定性处理；
- 基于设计和基本模型的不确定性评价；
- 随机空间模拟中的不确定性；
- 时空不确定性的处理；
- 分类数据的空间不确定性建模；
- 空间不确定性的可视化；
- 聚类和综合对空间不确定性评价的影响；
- 空间决定系统与不确定性的融合。

“ISSDQ2013”国际会议的主要内容包括：

- 现实世界实体的不确定性；
- 空间数据位置误差；
- 地理数据的时域不确定性；
- 空间数据的不完备性；
- 空间数据库的逻辑一致性；
- 空间数据的语义不确定性；
- 本体与空间数据质量；
- 空间数据综合的质量评价；
- 空间数据质量和尺度；
- GIS 中不确定性目标的空间数据模型；
- 不确定数据的空间查询；
- 不确定信息的空间推理；
- 空间数据质量的地学统计模型；
- 不完整的地面真实数据模型验证；
- 随机空间仿真；
- 遥感影像与分类的不确定性；
- DEM 的误差评定；
- 地理与环境分析的不确定性；
- 空间数据质量与决策；
- 空间数据质量与机器学习；

- 面向地理信息用户和公众的空间数据质量；
- 空间数据质量与基于 WEB 的服务；
- 众包空间数据质量评价；
- 地理可视化分析的不确定性；
- 空间数据可视化的质量；
- GIS 数据中元数据与模型；
- 基于设计和模型的空间数据质量评价；
- 多尺度与多源表达的不确定性；
- 空间数据质量管理；
- 网络与自发式数据质量评估；
- 多种应用的质量问题。

可见，国际上对地理信息的不确定性问题研究的内容越来越广泛，而且对于具体问题的研究越来越深入细致。

美国的 Kansas 大学、纽约州立大学 Buffalo 分校、Washington 大学、麻省理工学院（MIT）、Kent 州立大学、澳大利亚 Melbourne 大学、荷兰的 Amsterdam 大学等都设立了专门的空间数据的不确定性研究机构。我国的武汉大学、香港理工大学、山东科技大学、同济大学、中国科学院遥感应用研究所、北京大学遥感与地理信息系统研究所和中国矿业大学等单位都已经开始了这方面的研究。

国内外对空间数据不确定性的研究取得了长足进展，主要从 GIS 不确定性的研究内容与体系、数据不确定性的来源、数据不确定性的处理方法、空间分析过程中的不确定性传播机理等方面展开，目前，国内已有较为系统的研究并有专著出版，史文中（2005）出版了专著《空间数据与空间分析不确定性原理》；承继成、郭华东、史文中等人（2004）编著了《遥感数据的不确定性问题》；在王劲峰等人的著作《空间分析》一书中第五篇专门论述了空间信息的不确定性；邬伦等人（2010）出版了专著《地理信息系统中的不确定性问题》。国内外还有大量科技论文针对点、线、面的不确定性问题分别进行了研究，也有从位置、属性、时域和模型不确定性四个方面对空间数据的不确定性分别进行了研究。目前，关于地理

空间数据不确定性的研究现状如下：

①位置不确定性。位置不确定性包括绝对位置(坐标)不确定性和相对位置(空间关系)不确定性。绝对位置不确定性研究主要集中在对线元和面元不确定性模型的研究上，主要体现在对Chrisman(1982)提出的“ε-误差带”的扩展与改进以及置信域模型、可靠带模型等。点元的不确定性主要有随机模型和模糊模型，误差指标主要有误差区间(一维点)、误差椭圆(二维点)、误差椭球体(三维点)、超椭球体(多维点)、熵误差指标等。线元的不确定性以随机模型为主，主要模型有“ε-带”、“E-带”、“G-带”、纺锤形带。面元不确定性的随机模型误差指标为误差环。刘大杰等人利用“ε-误差带”模型，对GIS中点、直线、圆曲线、一般曲线、面的不确定性进行了研究，建立了误差带模型，导出了误差带的数学方程，分析了误差带的形状特征并加以可视化表达。邓敏、李志林、程涛(2006)还将误差带理论与粗集理论相结合，定义了点、线、面位置不确定性的上近似、下近似。

相对位置不确定性包括模糊空间拓扑关系、模糊方向关系等的研究。Cohn(1996)采用“鸡蛋-蛋黄”(egg-yolk)模型研究了带有不确定边界的空间区域，在RCC的理论框架下，基于蛋黄模型，用蛋和黄分别表示RCC中的区域外部和区域内部，最后根据RCC-5给出了不确定空间区域间的46种可能拓扑关系。Clementini等人(1998)在宽边界区域(regions with broad boundaries)的点集拓扑9-交模型的基础上研究了宽边界区域的9-交模型并对其进行了扩展，并采用代数方法得到了两个简单宽边界区域之间的44种拓扑关系。唐新明等人(2003)基于模糊数学和点集拓扑理论，研究了模糊区域的空间拓扑关系以及模糊空间中的空间对象建模问题。Beaubouef等人(2001)使用粗糙集描述了RCC-5和RCC-8。杜世宏等人(2003)用粗糙集研究了模糊空间对象间的空间方向关系，给出了方向关系粗糙表达方法、变精度粗糙表达方法以及方向关系的粗糙推理方法。他们结合模糊集和粗糙集给出了模糊对象的粗糙表达，基于点集拓扑的9-交模型理论，研究了模糊对象包括模糊面

与模糊面、模糊线与模糊面之间的拓扑关系；扩展后的9-交模型可以分辨简单模糊面之间的26种拓扑关系，还可以分辨模糊线和模糊面之间的38种拓扑关系。

②属性数据的不确定性。属性数据的不确定性研究相对较少，史文中，王树良(2001)在《GIS数据之属性不确定性的研究》一文详细介绍了属性不确定性的研究现状、内容和方法。

属性不确定性理论主要研究属性不确定性的来源、传播、位置与属性不确定性的关系。主要理论有模糊数学、粗集理论、证据理论和云理论等。主要的研究成果有高斯分布的分类数据不确定性自回归模型；基于知识推理的遥感数据解译不确定性模型；数据分类和边界划分不确定性模糊数学模型；从区域分类不确定性、边界定位误差和区域内部定量属性数据的抽样误差出发，综合进行属性数据不确定性的度量和传播分析；基于属性缺陷率模型的GIS数据质量控制等。

位置不确定性与属性不确定性在很多情况下往往是紧密相关、相互影响的。对此，史文中(1998)讨论了地理信息系统与遥感集成中位置与属性结合的不确定性问题，提出了“S-带”模型。

③时域不确定性。《空间数据现势度的概念》一文分析了空间数据现势性的不足之处，提出了空间数据现势度的概念，研究了现势度的四个性质及现势度和生命区间的影响因素。

④模型的不确定性。模型的不确定性不仅包括初始值和统计估计过程的不确定性，而且包括特殊模型形式和总体结构选择的不确定性，表现为选择不同的模型对同样的初始条件模拟，将得到不同的结果。可以依据模型可靠性原则和模型收敛原则，采用“总体平均可靠性评价”的方法，计算模型平均不确定性范围，从而评价各模型的不确定性；利用蒙特卡罗误差传播技术可以定量评价模型本身的不确定性和模型参数的不确定性等。

实际上上述各个方面是交叉影响的，空间的不确定性有时会影响到时域的不确定性，如多尺度的空间变量离散建模可以用于减少时态建模的不确定性等。目前，很多研究旨在把位置不确定性、属性不确定性、时域不确定性集成研究，希望找到一种集成表达

方法。

从目前的研究现状来看，GIS 不确定性的研究集中在对 GIS 基本实体点、线、面的不确定性的研究上。对于模型的不确定性，国外很多文献针对具体问题提出了一些评定和减少方法，但还未提出一种统一的模型不确定性评定方法。然而对 GIS 中空间分析、数据处理过程的不确定性，数据产品的不确定性以及 GIS 不确定性的内容体系的研究则较少。

邬伦(2010)根据现有空间数据不确定性研究议题及相关研究成果，提出了 GIS 数据不确定性框架体系，见表 1.3。

表 1.3　**GIS 数据不确定性框架体系及主要研究内容**

<table>
<tr><th colspan="2">GIS 数据的不确定性</th><th colspan="2">主要研究内容</th><th>主要研究方法</th></tr>
<tr><td rowspan="13">GIS 数据本身不确定性研究内容</td><td rowspan="11">位置数据的不确定性</td><td rowspan="3">数据源的不确定性</td><td>原始测量数据的不确定性(GPS、全站仪、数字测图成果的不确定性)</td><td rowspan="11">①经典误差理论以及在此基础上的空间统计理论与“ε-带”方法
②模糊集合理论
③灰色理论方法
④熵理论
⑤神经网络理论等方法</td></tr>
<tr><td>地图数字化(手扶跟踪数字化和扫描矢量化)</td></tr>
<tr><td>遥感数据的不确定性</td></tr>
<tr><td rowspan="4">GIS 中基本实体的不确定性</td><td>点实体的不确定性</td></tr>
<tr><td>线实体的不确定性</td></tr>
<tr><td>面实体的不确定性</td></tr>
<tr><td>体实体的不确定性</td></tr>
<tr><td rowspan="4">数据处理、空间分析过程的不确定性</td><td>数据格式转换的不确定性</td></tr>
<tr><td>缓冲区分析的不确定性</td></tr>
<tr><td>叠加分析的不确定性</td></tr>
<tr><td>其他分析的不确定性</td></tr>
<tr><td rowspan="2">属性数据的不确定性</td><td rowspan="2">属性定义的不确定性</td><td>对现实世界定义不明确</td><td rowspan="2">①目标模型和域模型
②概率论</td></tr>
<tr><td>模糊实体关系定义不确定</td></tr>
</table>

续表

<table>
<tr><th colspan="2">GIS数据的不确定性</th><th colspan="2">主要研究内容</th><th>主要研究方法</th></tr>
<tr><td rowspan="20"></td><td rowspan="5"></td><td rowspan="2">数据源的不确定性</td><td>实体边界界定的不确定性</td><td rowspan="5">③证据理论
④空间统计学
⑤模糊集合理论
⑥云理论
⑦粗集理论</td></tr>
<tr><td>遥感数据边界提取的不确定性</td></tr>
<tr><td rowspan="3">查询分析的不确定性</td><td>叠加分析的不确定性</td></tr>
<tr><td>查询过程中的不确定性</td></tr>
<tr><td>其他分析的不确定性</td></tr>
<tr><td rowspan="2">时域数据的不确定性</td><td rowspan="2">时间定义的不确定性</td><td>语义表达的模糊性带来的不确定性</td><td rowspan="5"></td></tr>
<tr><td>界定事件开始与结束的不确定性</td></tr>
<tr><td rowspan="3">逻辑的不确定性</td><td colspan="2">属性一致性的不确定性</td></tr>
<tr><td colspan="2">格式一致性的不确定性</td></tr>
<tr><td colspan="2">拓扑一致性的不确定性</td></tr>
<tr><td rowspan="2">数据的不完整性</td><td colspan="2">几何数据的不完整性</td><td rowspan="2"></td></tr>
<tr><td colspan="2">属性数据的不完整性</td></tr>
<tr><td rowspan="3">模型的不确定性</td><td colspan="2">模型选择的不确定性</td><td rowspan="3">①总体平均可靠性评价方法
②蒙特卡罗模拟</td></tr>
<tr><td colspan="2">模型参数的不确定性</td></tr>
<tr><td colspan="2">模型不确定性的传播与累加</td></tr>
<tr><td rowspan="5">不确定复合叠加</td><td colspan="2">属性数据不确定性对位置数据不确定性的影响</td><td rowspan="5">不确定性的叠加与传播</td></tr>
<tr><td colspan="2">时间变更对属性数据不确定性的影响</td></tr>
<tr><td colspan="2">时间变更对位置数据不确定性的影响</td></tr>
<tr><td colspan="2">位置数据与属性数据的不确定性叠加</td></tr>
<tr><td colspan="2">时空不确定性叠加</td></tr>
<tr><td rowspan="3">不确定性建模分析与表达</td><td>不确定性建模</td><td colspan="2">数据不确定值的数据模型建立</td><td rowspan="3">①科学计算可视化方法
②GIS可视化方法</td></tr>
<tr><td rowspan="2">数据不确定性可视化表达</td><td colspan="2">数据不确定性的显示与表达</td></tr>
<tr><td colspan="2">数据不确定性可视化结果分析</td></tr>
<tr><td rowspan="2">数据产品的不确定性评价</td><td>电子地图</td><td colspan="2">电子地图不确定性评价</td><td rowspan="2">不确定性的叠加与传播</td></tr>
<tr><td>4D产品</td><td colspan="2">DLG、DEM、DRG、DOQ产品的不确定性评价</td></tr>
</table>

续表

<table>
<tr><th colspan="2">GIS 数据的不确定性</th><th>主要研究内容</th><th>主要研究方法</th></tr>
<tr><td rowspan="5">GIS 工程不确定性评价与控制</td><td rowspan="3">GIS 组件最优（不确定性最小）分割</td><td>GIS 组件不确定性预测与评价</td><td rowspan="3">运筹学方法</td></tr>
<tr><td>GIS 组件间不确定性的叠加与传播</td></tr>
<tr><td>GIS 组件最优分割模型</td></tr>
<tr><td rowspan="2">GIS 工程不确定性估计与控制</td><td>GIS 工程不确定性预测与评价</td><td rowspan="2">建立预测模型软件工程的方法、项目管理方法</td></tr>
<tr><td>GIS 工程不确定性的控制方法与策略</td></tr>
</table>

空间数据不确定性的研究涉及测绘学、数学、地理学、计算机科学、地理信息科学等多门学科，是地学领域的基础理论和应用问题。空间数据不确定性的研究对评价空间数据质量、确定 GIS 数据录用标准、改善 GIS 算法、提高 GIS 分析结果的可信度、完善 GIS 基础理论和技术等有着重要的意义。

1.4.3　空间数据多尺度表达的研究进展

目前，有关多尺度现象的研究已经遍布各个学科分支，呈爆发式的趋势。但多尺度科学无论在国际还是在国内都还处于刚刚起步阶段。在地理信息科学领域，正掀起一股研究空间数据多尺度表达的热潮，国内外许多研究机构和高等学府都设有专门机构开展这方面的研究，如美国国家地理信息与分析中心（NCGIA）、法国国家地理研究院、瑞士苏黎世大学、英国爱丁堡大学、德国汉诺威大学、我国中科院地理科学与自然资源研究所、武汉大学、北京大学等。同时，还有一大批耳熟能详的著名学者在从事这方面的研究，如 Goodchild、Buttenfield、艾廷华、张景雄、李霖等。GIS 相关学术会议和空间论坛也将此作为中心议题，如空间数据处理国际研讨会（International Symposium on Spatial Data Handling）、GIS 帝国论坛（http：//www. gisempire. com/bbs）、地理信息科学国际会议（International Conference on Geographic Information Science）等。

相关理论与技术方法研究主要是从多尺度数据模型的形式化描

述、层次数据结构的设计和实现、不同层次之间的链接和更新触发机制以及不同表达版本的一致性维护等方面进行的。在概念模型及形式化描述方面，Puppo 等人（1995）引入抽象胞腔复形（abstract cell complexes）的概念，并建立了多尺度表达模型的形式化描述；基于空间认知的整体-部分构建原理，Stell 等人（1998）根据空间和语义精度概念提出了“层次地图空间（stratifiedmap spaces）”的概念，并将其用于多尺度变换过程（综合变换和类型转换）的形式化描述；同样是基于层次性原理，Timpf（1998）提出了地图立方体模型（map cube model），区分了 3 种不同的层次结构，即滤波层次、合并层次和综合层次；Jones 等人（1996）提出了一个基于知识推理（以知识或规则为基础进行推理）的多尺度概念模型——GEODYS-SEY；Balley 等人（2004）基于图章技术（stamping）扩展了 MADS 模型，并提供两种方法将其用于多尺度数据库的建立。在逻辑数据结构研究方面，考虑到空间数据多重表达之间固有的层次性，Oosterom（1995）提出了 GAP 和 BLG 数据模型，并将其分别用于多边形目标和线目标的层次化表达；Timpf（1995）研究了剥皮树和四叉树的局限性，引入并改进了图论中的 DAG（directed acyclic graph，有向非循环图）结构，将其用于描述多尺度表达中目标状态可能存在的出现/消失、合并/分裂等情况；为避免多版本存储所带来的数据冗余，Zhou 等人（2003）引入了多尺度几何体概念，将尺度和综合准则直接与每一坐标点挂钩，实现了同一几何体导出多重表达；针对特定的地理要素，如建筑物多边形，有关研究提出了操作性较强的逻辑数据结构，Sester 等人（2004）基于点的演变建立了建筑物目标多重表达的数据模型，Ai 等人（2004）则基于 LOD 层次化思想，提出了凸壳、矩形单元剖分后的变化累积模型，实现了不同精度下对建筑物多边形的拟合表达。

在开展理论与技术方法研究的同时，国内外还出现了一些空间数据多尺度表达的实用系统，如法国国家地理信息研究所建立了具有随尺度过渡关系的多重表示的实验系统；一些商用软件如 Smallworld GIS、Intergraph DynaGen 等基于初级尺度的自动综合方

法开发了部分多尺度表达方面的功能，Arc/Info（ArcView）、MapInfo 等采用重复建库的方法开发了多尺度表达方面的功能；VanOosterom 在 Postgres 中建立了一个交互式的多尺度 GIS 系统——GEO++；世界各国（美国、英国、澳大利亚、德国、日本、中国等）先后建立了多尺度空间数据基础设施，由于受多尺度处理技术的限制，大多存储的是具有固定比例尺的多个版本的数据；Internet 上流行的电子地图或影像查询网站，像 Google Map、Google Earth、Virtual Earth、GeoGlobe、MapPoint、GoeMap 等空间数据多尺度表达系统进一步解决了空间数据多尺度表达的理论和技术问题。

1.4.4　空间数据多尺度表达的不确定性研究进展

空间数据多尺度表达的相关问题是国内外的研究热点。大量空间数据多尺度表达的实用系统已被研发出来，且能生产出空间数据多尺度表达的产品。空间数据多尺度处理方法技术、产品的质量，即技术方法是否合理有效、产品能否满足用户的要求或者符合规范的要求，虽然有一些研究机构和学者已从事这方面的研究，但研究缺乏系统性。

空间数据多尺度表达是从空间数据衍生出来的，国内外对空间数据不确定性的研究取得了长足的进展。许多空间数据不确定性的研究与空间数据多尺度表达不确定性的研究相关，其研究成果可为空间数据多尺度表达不确定性研究利用和借鉴。综合整理空间数据不确定性和空间数据多尺度表达不确定性的研究成果，可看出对空间数据多尺度表达不确定性的研究主要分布在“多尺度空间数据的一致性、多尺度空间数据的相似性、多尺度空间数据的不确定性、多尺度空间数据的质量评价与控制”四个层面上。从广义上讲，这四个层面均可归纳为空间数据多尺度表达不确定性研究的内容。

在多尺度空间数据的一致性方面，评估不一致性的方法和维护策略是这一研究的重点内容，一致性维护必须保持每个目标的拓

扑、距离、方向和语义以及目标之间的空间关系。Egenhofer 等人（1994）提出了一个评估多重表示拓扑一致性的框架。Weibel（1996）针对多尺度空间数据库综合的不同内容的一致性约束进行了描述，并给出了相应的约束类型。李霖等人（2005）认为多尺度一致性约束是通过维护不同尺度上共同的空间结构特征实现的，并定义了一致性约束模型。Tryfona（1994）等人发展了评价拓扑关系等价性的形式化方法，并以此来评价多分辨率数据库中网络的拓扑等价性。杜晓初（2007）基于模糊函数量化了空间数据多尺度表达上的拓扑关系等价性规则和综合评价模型。目前，这方面的研究主要是从理论层面进行探讨的，还没有形成一种实用的评价方法或者模型。

在多尺度空间数据的相似性方面，主要是在数字环境下研究地理目标在不同尺度上的相似关系。代表性成果有 Goyal（2000）提出的一种能考虑空间目标在方向关系矩阵的方向片之间变化的空间方向相似性计算方法。Dimirtis Pdaias、Nikos Karacapilidis 和 Dinos Arkoumnais（1999）在定义了拓扑、方向和距离的相似程度后提出了一种空间场景相似程度的计算方法。H. T. Bunrs 和 M. Egenhoefr（1997）基于空间拓扑关系概念邻域图和空间方向关系概念邻域图提出了一种空间场景相似性的描述方法。Dimrtiis Papadias 等人（1999）用模糊隶属函数定义了空间拓扑关系的相似性、空间方向关系的相似性和空间距离关系的相似性。江浩等人（2009）研究并建立了多尺度地理空间点群目标相似关系的计算模型。张婷等人（2009）分析和探讨了多尺度下线状目标的相似因子。吕秀琴等人（2006）给出了空间对象拓扑关系相似性的定义以及在数字景观模型中拓扑关系相似性的度量方法。要解决多尺度空间数据生产结果的自动评价问题，必须系统研究多尺度的空间相似关系理论，而该问题的解决依赖于多尺度空间相似关系的定义、描述、模型设计、计算等一系列问题。当然，地理空间相似关系研究的意义远非空间数据多尺度表达所能包容的，它对于完善空间关系理论体系、空间查询与分析、空间推理等具有重要作用。目前，国内外对于地理空

间相似关系的研究成果基本处于空白，其研究过程非常复杂，需要数学、地图学、心理学等多学科联合攻关。

在多尺度空间数据的不确定性方面，McCarthy 等人（1956）早就指出：在地理研究中，不能期望在某一尺度上的研究得出的结论能适用于其他尺度，尺度的每一个变化都会反映出新问题，没有理由假设在某一尺度上的关联在其他尺度上仍然存在。Lam 等人（1992）将尺度和数据不确定性之间的关系归纳为四个方面：不同的量测单位其结果不一样；不同比例尺具有不同的不确定性；大尺度的研究覆盖较大的研究区域，代表性强，不确定性小；运行尺度越小把握性越大，不确定性越小。Openshaw（1977）系统地研究了地理学中的尺度效应。生态学家对生态尺度上体现出来的不确定性进行了大量研究，并针对尺度与生态结构关系提出了生态域和尺度门限的概念。针对多尺度遥感数据的不确定性问题的研究已有不少理论和实用技术，Woodcock 等人（1987）研究了遥感中尺度因子与遥感分类精度的关系，Niemann（1997）、Narayanan（2002）研究了遥感影像的分辨率对遥感图像信息内容的影响。徐丰等人（2006）针对线状要素多尺度表达的不确定性问题，从不确定性来源、传播机理、评价标准和模型等方面进行了研究，发展了适用于线状要素多尺度表达的不确定性的评估方法。吴凡（2005）发展了基于小波分析的尺度依赖的综合程度评估模型，并将其应用于多尺度表达的线状、面状要素中。

在多尺度空间数据的质量评价与控制方面，主要是用设计和试验评估综合过程与质量的方法。IGN 在开发实用综合程序的过程中，区别对待地理信息的多级比例尺表示（图形）和多级分辨率显示（图像），并对这两种不同性质的综合问题规定了不同的质量控制指标。Buttenfield、McMaster 和 Plazanet 应用多个应用数学或几何方法完成了综合质量评估的研究报告（Joao，1998），徐丰等人（2009）研究了线状要素多尺度表达质量评价的概念模型。钱海忠等人（2005）提出一种基于极化变换的点群综合几何质量评估方法。然而，目前还没有关于这些已有的和潜在的评估方法及其

特性的综合性框架。

1.5 本章小结

本章首先对空间数据多尺度表达所涉及的相关名词的概念进行了分析和阐述，这些相关名词包括：空间数据、尺度、空间数据多尺度表达、不确定性等；其次，在阐述空间数据不确定性的基础上，探讨了空间数据多尺度表达不确定性的存在性，并对空间数据多尺度表达不确定性进行了定义，在这里将空间数据多尺度表达不确定性定义为多种比例尺表达的空间数据与空间现象本身固有的属性之间的差异；然后，从四个方面阐述了研究空间数据多尺度表达不确定性的理论与实践意义；最后，阐述了空间数据多尺度表达不确定性及其相关研究的研究进展，包括不确定性的认知历程、地理空间数据不确定性的研究进展、空间数据多尺度表达的研究进展、空间数据多尺度表达不确定性的研究进展。

第2章　空间数据多尺度表达不确定性的理论基础

现实世界的不确定性问题可归纳为现实世界客观存在的不确定性和人们主观认识上的不确定性两类。现实世界客观存在的不确定性最为典型的是随机性；人们主观认识上的不确定性有：由于定义不清造成的模糊性，灰色系统中的部分已知、部分未知的灰性，以及未确知性，未确知性是指就客观世界本身来讲它是确定的，但对于决策者来说由于受决策条件限制而导致的认识不清，所掌握的证据不足以确定事物的真实状态和数量关系而带来的纯主观的、认识上的不确定性。这四类不确定性中，随机不确定性和模糊不确定性是强不确定性，灰性和未确知性是弱不确定性。

本章首先对本书处理现实世界随机不确定性和模糊不确定性所涉及的几种实用而且常见的数学理论进行阐述，然后对已有的空间数据随机不确定性和模糊不确定性模型进行论述和分析。这些数学理论与不确定性模型是构成本书研究空间数据多尺度表达不确定性的理论基础。

2.1　数学基础

2.1.1　概率论与数理统计

概率论与数理统计是研究随机现象的统计规律的一门学科。概率论是在已知随机变量服从某种分布的情况下研究随机变量分布（分布函数、分布律、密度函数）的性质和数字特征（如数学期望

值、方差、相关系数等）的性质及其应用。数理统计学是以概率论为基础，研究如何以有效的方式收集、整理和分析受到随机性影响的数据，并以这些有限的数据为依据对所考察的问题作出判断和预测，为采取某种决策提供依据。它主要包括以下两个方面：

试验设计——研究如何对随机现象进行观察、试验，以取得有代表性的局部观察值。

统计推断——研究如何对试验获得的数据进行整理、分析，从而对所关心的问题作出尽可能精确、可靠的结论，即对随机变量的分析和分布中的参数进行估计和判断。

由于概率论与数理统计是使用局部现象去推断整体内在规律性的方法，在现代科学中具有典型意义，应用极其广泛。

1. 随机事件与概率

在一定条件下，随机试验可能发生也可能不发生的结果称为随机事件，简称事件。最简单的不可再分的事件称为样本点或基本事件，由随机试验所有可能结果组成的集合称为样本空间，记作 Ω。样本空间 Ω 的一个特殊子集为 Ω 本身，称为必然事件；另一个称为不可能事件，记作 $\varnothing$。

1）频率与概率

概率的统计定义为：在一定条件下，重复做 n 次试验，事件 A 发生的频数为 m，则事件 A 发生的频率为 $\frac{m}{n}$；若 n 足够大，$\frac{m}{n}$ 会趋向于某一常数值 p，即 $\lim\limits_{n\to\infty}\frac{m}{n}=p$，则称常数 p 为事件 A 的概率，记为 $P(A)=p$。

显然，数值 p 是事件 A 发生的可能性大小的客观数量描述，$0\leqslant P(A)\leqslant 1$，样本空间的概率 $P(\Omega)=1$，不可能事件的概率 $P(\varnothing)=0$。概率的统计定义提供了一种概率值的近似计算方法，即采用大量试验事件中的频率 $\frac{m}{n}$ 作为事件概率 p 的近似值。一般情况下，n 越大，近似程度越高。

2）概率的计算公式

（1）条件概率与乘法公式

若 A 和 B 为两个随机事件，在事件 A 已经发生的条件下，事件 B 发生的概率称为在事件 A 发生的条件下，事件 B 的条件概率，或称为事件 B 关于事件 A 的条件概率，记为 $P(B\mid A)$。当 $P(A)=0$ 时，规定 $P(B\mid A)=0$，并得乘法公式：

$$P(AB)=P(B\mid A)P(A)=P(A\mid B)P(B)$$

$P(A_1A_2\cdots A_n)=P(A_n\mid A_1A_2\cdots A_{n-1})P(A_{n-1}\mid A_1A_2\cdots A_{n-2})\cdots P(A_2\mid A_1)P(A_1)$。

(2) 独立性公式

如果事件 A 和 B 满足 $P(A\mid B)=P(A)$，则称事件 A 关于事件 B 是独立的。事件 A 和 B 独立的充要条件是：

$$P(AB)=P(A)\cdot P(B)$$

(3) 全概率公式

设事件组 $\{B_i:i\in\mathbf{I}\}$ 为样本空间 Ω 的一个划分，且 $P(B_i)>0$ $(i\in\mathbf{I})$，$\mathbf{I}=1,2,\cdots,n$。

则对任一事件 A 有：

$$P(A)=\sum_{i\in\mathbf{I}}P(A\mid B_i)P(B_i)$$

上面的公式称为全概率公式。在使用全概率公式解决实际问题时，关键在于构建一个完备事件组。

(4) 贝叶斯公式

如果一个事件组 $\{B_i:i\in\mathbf{I}\}$ 为样本空间 Ω 的一个划分，且 $P(A)>0$，$P(B_i)>0(i\in\mathbf{I})$，则对于任何 $i\in\mathbf{I}$，有

$$P(B_i\mid A)=\frac{P(B_i)P(A\mid B_i)}{\sum_{i\in\mathbf{I}}P(B_i)P(A\mid B_i)}$$

2. 随机变量与分布函数

(1) 随机变量与概率分布函数的定义

每次试验的结果可以用一个变量 X 的数值来表示，这个变量的取值随偶然因素而变化，但遵从一定的概率分布规律，这种变量称为随机变量。

给定随机变量 X，它的取值不超过实数 x 的事件的概率 $P(X\leqslant x)$ 是 x 的函数，称为 X 的概率分布函数，简称分布函数，记作

$F(x)$，即

$$F(x)=P(X\leqslant x)(-\infty<x<+\infty)$$

（2）随机变量的函数的分布

如果随机变量 y 是随机变量 x 的函数 $y=f(x)$，记随机变量 x 的分布函数 $F(x)$ 为 $F(x)=\int_{-\infty}^{x}\mathrm{d}f(y)$。特别地，当 x 是离散型随机变量时，其可能值为 $x_1,x_2,\cdots$，且 $P(X=x_k)=p_k(k=1,2,\cdots,n)$，则

$$F(x)=\sum_{f(x_k)\leqslant x}p_k$$

当 x 是连续型随机变量时，其分布密度为 $f(x)$，则

$$F(x)=\int_{-\infty}^{x}f(y)\mathrm{d}y$$

3. 随机变量的数字特征

（1）数学期望与方差

随机变量 X 的数学期望值记作 $E(X)$，它描述了随机变量的取值中心。$[X-E(X)]^2$ 的数学期望值称为随机变量 X 的方差，记作 $D(X)$，$D(X)$ 的平方根称为随机变量 X 的标准差或均方差，记作 $\sigma_X=\sqrt{D(X)}$。方差和均方差描述了随机变量的可能取值与数学期望的偏离程度。

若 X 是连续型随机变量，其分布密度函数为 $f(x)$，分布函数为 $F(x)$，则

$$E(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x=\int_{-\infty}^{\infty}x\mathrm{d}F(x)$$

$$D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2f(x)\mathrm{d}x=\int_{-\infty}^{\infty}[x-E(X)]^2\mathrm{d}F(x)$$

若 X 是离散型随机变量，其可能取值为 $x_k,k=1,2,\cdots,n$，且 $P(X=x_k)=p_k$，则

$$E(X)=\sum_{k=1}^{\infty}x_k p_k$$

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k$$

若 $X_1,X_2,\cdots,X_n$ 为 n 个随机变量，则

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$$

$$D(X_1+X_2+\cdots+X_n)=\sum_{i,j=1}^{n}E[(X_i-E(X_i))(X_j-E(X_j))]$$

若$X_1,X_2,\cdots,X_n$ 为 n 个相互独立的随机变量,则

$$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots(X_n)$$

若$X_1,X_2,\cdots,X_n$ 为 n 个相互独立的随机变量, 且 $E(X_k)=0$, $D(X_k)=\sigma^2(k=1,2,\cdots,n)$,则随机变量 $Y=\frac{1}{n}\sum_{k=1}^{n}X_k$ 的数学期望与方差分别为:

$$E(Y)=0,D(Y)=\frac{\sigma^2}{n}$$

(2) 协方差与相关系数

设随机变量X和Y的数学期望与方差都存在,则X和Y的协方差σ_{XY} 或 $\mathrm{Cov}(X,Y)$ 为:

$$\sigma_{XY}=E[(X-E(X))(Y-E(Y))]$$

X 和 Y 的相关系数ρ_{XY} 为:

$$\rho_{XY}=\frac{E[(X-E(X))(Y-E(Y))]}{\sqrt{D(X)}\sqrt{D(Y)}}$$

4. 总体参数的估计

1) 总体与样体

通常把被研究的对象的全体称为总体(或母体),记作X,把组成总体的元素称为个体, 记作 x。从总体中随机抽取 n 个个体x_1, $x_2,\cdots,x_n$ 为总体的一个样本(或子样)。样本中样品的个数称为样本的大小(或容量)。一般情况下,$n>30$ 可以认为是大样本,否则称为小样本。

2) 总体参数的估计

记$x_1,x_2,\cdots,x_n$ 是从总体 X 中取出的一个样本,可用样本的特征数来估计总体的数字特征,主要方法有矩法和极大似然法。

(1) 矩法

矩法是用样本的 r 阶矩作为总体 r 阶矩的估值, 具体实现途径为:

设总体 X 的分布函数为 $F(x;\theta_1,\theta_2,\cdots,\theta_k)$,即有 k 个未知参数

$\theta_1,\theta_2,\cdots,\theta_k$。假定 X 的 k 阶原点矩存在，它们是$\theta_1,\theta_2,\cdots,\theta_k$ 的函数，即

$$\upsilon_r = \upsilon_r(\theta_1,\theta_2,\cdots,\theta_k) = \int_{-\infty}^{\infty} x^r \mathrm{d}F(x;\theta_1,\theta_2,\cdots,\theta_k)\ (r = 1,2,\cdots,k)。$$

考虑总体的一个样本$x_1,x_2,\cdots,x_n$，作出这一样本的 r 阶矩$\hat{\upsilon}_r$，即

$$\hat{\upsilon}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^k \quad (r = 1,2,\cdots,k)$$

记所得的解为

$$\hat{\theta}_1 = \hat{\theta}_1(x_1,x_2,\cdots,x_n),\cdots,\hat{\theta}_k = \hat{\theta}_k(x_1,x_2,\cdots,x_n)$$

用$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 分别作为$\theta_1,\theta_2,\cdots,\theta_k$ 的估值。

(2) 极大似然法

设总体的分布是连续型的，分布密度函数为$p(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中$\theta_1,\theta_2,\cdots,\theta_k$ 是待估计的未知参数，对于给定的$x_1,x_2,\cdots,x_n$，使函数$\prod_{i=1}^{n} p(x_i;\theta_1,\theta_2,\cdots,\theta_k)$ 达到最大值的$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$，并用它们作为$\theta_1,\theta_2,\cdots,\theta_k$ 的估值。

由于 $\ln\prod_{i=1}^{n} p(x_i;\theta_1,\theta_2,\cdots,\theta_k)$ 和 $\prod_{i=1}^{n} p(x_i;\theta_1,\theta_2,\cdots,\theta_k)$ 在同一点$(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)$ 上达到最大值，因此，引入似然函数

$$\begin{aligned} L(\theta;\theta_1,\theta_2,\cdots,\theta_k) &= \ln\prod_{i=1}^{n} p(x_i;\theta_1,\theta_2,\cdots,\theta_k) \\ &= \sum_{i=1}^{n} \ln p(x_i;\theta_1,\theta_2,\cdots,\theta_k) \end{aligned}$$

只要解方程组

$$\frac{\partial L}{\partial \theta_i} = 0 \quad (i = 1,2,\cdots,k)$$

就可以求出$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$，它们称为$\theta_1,\theta_2,\cdots,\theta_k$ 的最大似然估计值。

对于总体分布为离散的形式，只需把 $p(x;\theta_1,\theta_2,\cdots,\theta_k)$ 改为 $P(X = x_i)$ 即可。

(3) 估值的判别标准

估计值好坏的判别标准有三：

① 无偏性：$\hat{\theta}$ 与被估计参数 θ 的真值越接近越好。若估计值 $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ 满足关系式 $E(\hat{\theta}_n)=\theta$，则称 $\hat{\theta}_n$ 是 θ 的无偏估计量。

② 有效性：$\hat{\theta}$ 围绕被估计参数 θ 的真值波动的幅度越小越好。如果 $\hat{\theta}$ 和 $\hat{\theta}'$ 都是参数 θ 的无偏估计，如果 $D(\hat{\theta}) \leqslant D(\hat{\theta}')$，则称 $\hat{\theta}$ 比 $\hat{\theta}'$ 有效。

③ 一致性：当样本容量越来越大时，$\hat{\theta}$ 靠近被估计参数 θ 真值的可能性也应该越来越大。最好是当样本容量趋于无穷时，$\hat{\theta}$ 在某种意义上收敛于 θ 的真值。若对任意给定的正数 ε，总有 $\lim\limits_{n\to\infty}P(|\hat{\theta}_n-\theta|>\varepsilon)=0$，则称 $\hat{\theta}_n$ 是总体参数 θ 的一致估计量。

5. 假设检验

1）假设检验的一般步骤

先假设总体具有某种统计特性（如具有某种参数或服从某种分布等），然后检验这个假设是否可信，这种方法称为假设检验（或统计假设检验）。其步骤如下：

① 根据问题的实际情况，建立原假设 H_0 及备选假设 H_1；

② 选定检验统计量并分析拒绝域的形式；

③ 给出显著性水平 α，并由此确定拒绝域 C；

④ 取样，根据样本观察值做出判断是否拒绝 H_0。

2）总体分布函数的假设检验

设 $F_0(x)=F_0(x;\theta_1,\theta_2,\cdots,\theta_k)$，其分布函数的类型已知，参数 $\theta_1,\theta_2,\cdots,\theta_k$ 已知或部分已知，总体 X 的样本为 $x_1,x_2,\cdots,x_n$，总体 X 的分布函数为 $F(x)$，其假设检验分如下两种情况进行：

（1）$F_0(x)$ 的全部参数已知

将实轴 $(-\infty,+\infty)$ 划分成 m 个不相交的区间：$(c_i,c_{i+1}]$ $(i=1,2,\cdots,m)$，$c_1=-\infty$，$c_{m+1}=+\infty$。记理论频率为

$$r_i=F_0(c_{i+1})-F_0(c_i)=P(c_i<X\leqslant c_{i+1})$$

X 的样本 $x_1,x_2,\cdots,x_n$ 落在区间 $(c_i,c_{i+1}]$ 的个数为 v_i（经验频

数），那么统计量$\chi^2=\sum_{i=1}^{m}\frac{(v_i-nr_i)^2}{nr_i}$服从自由度为$m-1$的$\chi^2$分布，应用$\chi^2$检验法可检验假设$H_0:F_0(x)=F_0(x)$是否可信。

（2）$F_0(x)$的参数全部或一部分未知

设$F_0(x)$有l个参数$\theta_{ji},\theta_{j2},\cdots,\theta_{jl}(l\leqslant k)$未知，可先用最大似然法给出这$l$个参数的估值，把这些估值就当作$F_0(x)$的相应参数。于是类似于(1)，可计算理论频率，再计算经验频数，那么当n很大时统计量$\chi^2=\sum_{i=1}^{m}\frac{(v_i-nr_i)^2}{nr_i}$服从自由度为$m-l-1$的$\chi^2$分布，应用$\chi^2$检验法可检验假设$H_0:F_0(x)=F_0(x)$是否可信。

2.1.2　信息论与熵

信息论来源于通信领域，是由Shannon和Weaver在概率论的基础上提出来的，是以不确定性来表示信息，如有信息存在，则就有不确定性存在，信息的传递就是消除不确定性；若没有不确定性，就没有信息（Shannon等，1959）。目前，信息理论已经成功地应用到图像处理和模式识别中，也开始应用到地图传输和GIS中。Blais（1991）指出，在空间信息处理，特别是在量化含糊、模糊等不同类型的不确定性处理方面，信息理论具有很大的潜力。

熵的主要含义是指当认识的主体对客体缺乏必要的知识时，往往表现为对被研究的客体具有某种不确定性，当获得信息后，这种不确定性可以减少或消除。主体获得信息的多少，与客体不确定性的消除程度有关。熵在空间数据不确定性研究中的应用首先表现在对测量结果随机性的度量（胡圣武，2006）。

1. 信源的数学模型

信源就是信息的载体。信源要含有一定的信息，随机性地以一定的概率发出各种不同的信号。单符号离散信源最为简单，它是具有一定概率分布的离散符号的集合，因此可用一个离散随机变量来代表它。

如果一个信源可能发出r种不同的符号$a_1,a_2,\cdots,a_r$，设其相应的先验概率分别为$p(a_1),p(a_2),\cdots,p(a_r)$，若用随机变量$X$表示此

信源,其信源空间可表达为

$$[X \cdot P]:\begin{cases} X:a_1 & a_2 & \cdots & a_r \\ P(X):p(a_1) & p(a_2) & \cdots & p(a_r) \end{cases}$$

上式中,$0 \leqslant p(a_i) \leqslant 1(i=1,2,\cdots,r)$,$\sum_{i=1}^{r} p(a_i)=1$。

信源空间与信源是一一对应的关系,不同信源空间对应于不同的信源,不同的信源也对应着不同的信源空间。

2. 自信量

根据信息的定义可知,在通信过程中,收信者获得的信息量等于通信前后不确定性的消除。若信源发出某种信号是a_i,由于信道中存在噪声,使得收信者收到的是a_i的改变型b_j,收信者从信号b_j获取原始信号a_i的信息量$I(a_i;b_j)$就为:

$I(a_i;b_j)$=[收到b_j前,收信者对信号a_i的不确定性]-[收到b_j后,收信者对信号a_i仍然存在的不确定性]=收信者收到b_j前后,对信号a_i的不确定性的消除

当信道中没有噪声的干扰时,信源发出的信号a_i可以准确地传递给收信者,也就是说,$b_j=a_i$。这等同于对信源发信号a_i的不确定性为0,此时,$I(a_i;b_j)$可简化为:

$I(a_i)=I(a_i;a_i)$=[收到a_i前,收信者对信源发a_i的不确定性]

上式表明,对信源信号a_i的自信量$I(a_i)$的度量问题已转变为信源发信号a_i的不确定性的度量问题。因此,自信量$I(a_i)$是信源发信号a_i的先验概率$p(a_i)$的某一函数,即

$$I(a_i)=f[p(a_i)](i=1,2,\cdots,r)$$

而且,函数式$I(a_i)$是$p(a_i)$的倒数的对数,即

$$I(a_i)=\log\frac{1}{p(a_i)}=-\log p(a_i)(i=1,2,\cdots,r) \tag{2.1}$$

由式(2.1)表达的函数可以作为信源

$$[X \cdot P]:\begin{cases} X:a_1 & a_2 & \cdots & a_r \\ P(X):p(a_1) & p(a_2) & \cdots & p(a_r) \end{cases}$$

发信号$a_i(i=1,2,\cdots,r)$的不确定性。式(2.1)被称为信息函数。由信息函数可知,信源发信号$a_i(i=1,2,\cdots,r)$的不确定性由信号

$a_i(i=1,2,\cdots,r)$ 的先验概率$p(a_i)(i=1,2,\cdots,r)$ 唯一确定。从这个意义上说，式(2.1) 又称为概率信息。从式(2.1) 可知，自信息量$I(a_i)$ 的单位取决于对数的“底”，若以 2 为底，则所得自信息量的单位为“比特”，即

$$I(a_i)=\log_2\frac{1}{p(a_i)}\text{ 比特}$$

3. 信息熵

信息熵是整个信源的总体信息测度。信息熵 $H(X)$ 定义为：信源 X 发出的各种不同信号$a_i(i=1,2,\cdots,r)$ 含有的自信息量$I(a_i)$ $(i=1,2,\cdots,r)$ 在信源的概率空间，$\{p(a_1),p(a_2),\cdots,p(a_r)\}$ 中的统计平均值，即

$$\begin{aligned}H(X)&=p(a_1)I(a_1)+p(a_2)I(a_2)+\cdots+p(a_r)I(a_r)\\&=-\sum_{i=1}^{r}p(a_i)\log p(a_i)\end{aligned}$$

它表示信源 X 每发一个信号所提供的平均信息量。

信息函数 $I(a_i)$ 既表示收信者确切无误收到信源信号a_i 后，从a_i 中获取的信息量，同时也表示收到信号a_i 前，收信者对于信源发信号a_i 存在的不确定性。信息熵$H(X)$，除表示信源X每发一个信号所提供的平均信息量之外，也表示收信者在收到信号前，对信源X存在的平均不确定程度。

令$p_i=p(a_i)$，信息熵改写为：

$$H(X)=-\sum_{i=1}^{r}p_i\log p_i$$

由于 $H(X)$ 为p_i 的函数，因此，$H(X)$ 可进一步表示为：

$$H(p_1,p_2,\cdots,p_r)=-\sum_{i=1}^{r}p_i\log p_i$$

若把信源X的r个信号$a_i(i=1,2,\cdots,r)$ 的概率分布$p_1,p_2,\cdots,p_r$ 看作概率矢量 P 的 r 个分量，即

$$P=(p_1,p_2,\cdots,p_r)$$

则熵函数又可表示为：

$$H(P)=H(p_1,p_2,\cdots,p_r)$$

设信源 X 的信源空间为：

$$[X \cdot P]:\begin{cases} X: a_1 \quad a_2 \quad \cdots \quad a_r \\ P(X): p(a_1) \quad p(a_2) \quad \cdots \quad p(a_r) \end{cases}$$

概率分量也等于 r 的另一完备的概率空间 $Q:\{q_1,q_2,\cdots,q_r\}$，且

$$H(p_1,p_2,\cdots,p_r) \leqslant -\sum_{i=1}^{r} p_i \log q_i$$

这就是熵函数的极值性，它表明：信号种数为 r 的信源 X 的熵函数，一定不大于概率分数量同样是 r 的另一信源中每一信号的自信息量，在信源 X 的概率空间中的统计平均值。这意味着，熵函数存在最大值。

2.1.3　模糊理论

现实世界充满着各种模糊对象(事物、现象等)，到处是各种模糊联系(模糊空间关系等)。根据哲学反映论的观点，人们对世界的认识，亦即人的认知，也必然是充满模糊性的。这就决定了人们不可能完全采用精确方法来描述和表达现实世界，而必须采用模糊的方法来研究和表达客观世界。要处理模糊世界必然要面对其中的各种模糊对象和模糊概念等，对这些模糊概念或对象的处理不能再套用那些精确知识的处理方法，必须要结合模糊性研究中的各种特殊的处理方法和技术。要求其一是它处理的对象或概念是模糊的；其二是这种处理方法本身应该包含模糊性，即在描述、处理、分析与应用中所做的操作和运算本身具有某种模糊性，称为模糊建模。模糊集合理论被认为是模糊建模最有力的理论工具，因此得到了广泛的关注。

1. 模糊集合

一个概念的外延如果是不明确的、边界不清，这个概念就是一个模糊概念，或者说一个概念对应的集合是一个模糊集合。在经典数学中，集合可用特征函数来表示，论域 U 中的元素 $\forall x \in U$ 属于普通集合 P，则特征函数$X_p(x)=1$，否则$X_p(x)=0$。对于模糊集合来说，U 中存在不同程度属于一模糊集的元素，如果利用特征函数来表达模糊集，则模糊集对应的特征函数的取值就不再是$\{0,1\}$，而是在整个闭区间$[0,1]$ 中取值。Zadeh 教授的贡献之一就是基于此提出了模

糊集合的概念。

(1) 模糊集合定义

Zadeh(1965)基于以上思想,给出了模糊集合的概念:

论域 U 上的模糊子集 A 是指对于任意 $x \in U$,都有一个数$\mu_A(x) \in [0,1]$ 与之对应,即有映射:

$$\mu_A: U \to [0,1]; x \to \mu_A(x)$$

映射$\mu_A(x)$ 成为模糊子集 A 的隶属函数,$\mu_A(x)$ 表示元素 x 隶属于 A 的程度。模糊子集由其隶属函数所确定,隶属函数是对普通集合特征函数的推广。

(2)λ 截集

设 $A \in F_{(u)}$,对于任意 $\lambda \in [0,1]$,称$A_\lambda = \{u \mid u \in U, \mu_A(u) \geq \lambda\}$ 为模糊集合 A 的 λ 截集,λ 称为置信水平,如果定义式中不等号"$\geq$"换成"$>$",则称这样的 λ 截集为 λ 强截集或开截集。

λ 截集为一普通集合,其对应的特征函数为:

$$X_{A\lambda}(u) = \begin{cases} 0, 当\mu_A(u) < \lambda; \\ 1, 当\mu_A(u) \geq \lambda, \end{cases}$$

(3) 模糊集合的贴近度

设映射 $\sigma: \mathcal{F}(U) \times \mathcal{F}(U) \mapsto [0,1]; (A,B) \mapsto \sigma(A,B) \in [0,1]$,并且如果 σ 满足下列条件:

①$\sigma(A,B) = \sigma(B,A)$;

② 当 $A = B$,有 $\sigma(A,B) = 1$;

③$\sigma(\varnothing, U) = 0$;

④ 若 $A \subseteq B \subseteq C$,则 $\sigma(A,C) \leq \sigma(A,B) \wedge \sigma(B,C)$,

则称映射σ 为U的模糊集合的贴近度。贴近度反映了模糊集合的接近程度,贴近度有不同的计算方法:

① 基于距离的贴近度计算:

$$\sigma(A,B) \triangleq 1 - c[d_p(A,B)]^\alpha$$

其中,$d_p(A,B) \triangleq \left(\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^p\right)^{\frac{1}{p}}$。当$p \geq 1$时,称$\sigma$ 为 A 与 B 之间的闵可夫斯基(Minkowski)距离;当$p = 1$ 时,该距离成为海明(Hamming)距离;当 $p = 2$ 时,相应的距离成为欧几里得

(Euclid) 距离。

② 基于取大、取小运算的贴近度计算：

$$\sigma(A,B) \triangleq \frac{\sum_{i=1}^{n}(\mu_A(x_i) \wedge \mu_B(x_i))}{\sum_{i=1}^{n}(\mu_A(x_i) \vee \mu_B(x_i))}$$

或
$$\sigma(A,B) \triangleq \frac{2\sum_{i=1}^{n}(\mu_A(x_i) \wedge \mu_B(x_i))}{\sum_{i=1}^{n}(\mu_A(x_i) + \mu_B(x_i))}$$

2. 模糊隶属函数的确定方法

模糊隶属函数是用模糊理论处理问题的关键，关于模糊隶属函数的确定目前还没有形成通用的方法，往往需要根据实际应用采取不同的策略确定，以下介绍几种常用的方法：

(1) 模糊统计方法

该方法类似于随机试验方法，其基本原理是在进行模糊统计试验时，发现统计结果呈现一定的统计规律——隶属频率的稳定性。

设 $A \in F(U)$，$u_0 \in U$，通过模糊统计实验来确定 u_0 对 A 的隶属度，即 $\mu_A(u_0)$，若做 n 次这样的模糊试验，其中 m 次 A^*（A^* 是由每次试验判断结果构成的一个边界可变的，可以移动的普通集合），则称 $\frac{m}{n}$ 为 u_0 对 A 的隶属频率 t，即

$$t = \frac{m}{n} = \frac{u_0 \in A^* \text{ 的次数}}{\text{试验总次数}}$$

随着 n 增大，隶属频率会出现稳定性，而它稳定所在的那个值便定为 u_0 对 A 的隶属度 $\mu_A(u_0)$。对论域中的每一个元素不断运用上述方法，最终会获得论域 U 上模糊子集 A 的隶属函数。

该方法的关键是模糊统计试验，归纳起来，模糊统计试验包括四个要素：论域 U；U 中的一个元素 u_0；U 中一个边界可变的普通集合 A^*，它体现了模糊概念之外的外延的不确定性，即边界不确定；条件，它体现了模糊概念 A 影响之下的划分过程，并制约着 A^* 的运动。

（2）二元对比排序法

这种方法是将论域中的每一个元素同其他元素进行充分的两两比较，属于模糊集 A 的程度后，经过处理得出各自的隶属度。

设论域 $U=\{u_1,u_2,\cdots,u_n\}$，$A\in F(U)$。$\forall\ u_i,u_j\in U$，用r_{ij} 表示 u_i 关于 A 比u_j 优先的程度，并做如下限定：

① $0\leqslant r_{ij}\leqslant 1$，$(i,j=1,2,\cdots,n)$；

② $r_{ii}=1$，$(i=1,2,\cdots,n)$；

③ $r_{ij}+r_{ji}=1$，$(i,j=1,2,\cdots,n)$；

将r_{ij} 作为元素的一个方阵，然后根据实际情况分别作出如下不同处理：

取小处理，矩阵的第 i 行 $(i=1,2,\cdots,n)$ 反映出u_i 对 A 的隶属度，即$\mu_A(u_i)=\wedge_{j=1}^{n} r_{ij}(i=1,2,\cdots,n)$，最后可得到

$$A=\{\mu_A(u_1),\mu_A(u_2),\cdots,\mu_A(u_n)\}$$

或者作算术平均处理（或加权平均），即将矩阵中每一行求平均值，以此作为各元素的隶属度，即$\mu_A(u_i)=\dfrac{1}{n}\sum\limits_{j=1}^{n} r_{ij}$或$\mu_A(u_i)=\dfrac{1}{n}\sum\limits_{j=1}^{n}\delta_j r_{ij}$，其中$\sum\limits_{j=1}^{n}\delta_j=1$。

2.1.4 粗集理论

1982 年，波兰学者 Z. Pawlak 提出了粗集理论——它是一种刻画不完整性和不确定性的数学工具，能有效地分析不精确、不一致（inconsistent）、不完整（incomplete）等各种不完备的信息，还可以对数据进行分析和推理，从中发现隐含的知识，揭示潜在的规律。粗集理论是建立在分类机制的基础上的，它将分类理解为在特定空间上的等价关系，而等价关系构成了对该空间的划分。该理论将知识理解为对数据的划分，每一被划分的集合称为概念，它的主要思想是利用已知的知识库，将不精确或不确定的知识用已知的知识库中的知识来（近似）刻画。

近年来，一些学者运用粗集理论对 GIS 数据的不确定性进行分析建模。

1. 粗集的基本概念

设 R 为论域 U 上的一组等价关系,并将 U 划分为互不相交的基本等价类,二元对 $K=(U,R)$ 构成一个知识库(或近似空间),$(U,R)=\{X_1,X_2,\cdots,X_n\}$。对于一个给定的集合 X,满足 $X\subseteq U$,w 为 U 中的一个对象,$[w_R]$ 表示所有与 w 不可分辨的对象组成的集合,即由 w 决定的等价类。当集合 X 能表示成基本等价类组成的并集时,则称集合 X 是可精确定义的;否则,集合 X 只能通过逼近的方式来刻画。在粗集中,这种逼近是通过集合 X 关于 R 的下近似和上近似来实现的。它们可分别表达为:

集合 X 关于 R 的下近似:$\underline{R}(X)=\{w\in U:[w]_R\subseteq X\}$

集合 X 关于 R 的上近似:$\overline{R}(X)=\{w\in U:[w]_R\cap X\neq\varnothing\}$

集合 X 的边界区:$\mathrm{bn}(X)=\overline{R}(X)-\underline{R}(X)$

2. 粗糙度与分类质量

粗集的不确定性主要是由两个原因引起的:一个原因是来自于给定近似空间的粗集的边界,这种不确定性称为系统的不确定性;另一个原因直接来自于论域上二元关系对论域中对象进行分类的能力,即知识的粒度。知识粒度越大,相对于近似空间中的概念和知识就越不确定,这种不确定性称为概念的不确定性。

粗集理论处理系统的不确定性是通过引进近似精度与粗糙度来实现的,它们可分别表达为:

集合 X 的近似精度:$\alpha_R(X)=\dfrac{\underline{R}(X)}{\overline{R}(X)}$

集合 X 的粗糙度:$\rho_R(X)=1-\alpha_R(X)$

集合 X 的近似精度$\alpha_R(X)$ 和粗糙度$\rho_R(X)$ 均是一个区间$[0,1]$上的实数,它们分别定义了集合 X 的可定义程度和知识的不完全程度。$\alpha_R(X)$ 越大,集合 X 的确定度越大;与近似精度$\alpha_R(X)$ 恰恰相反,$\rho_R(X)$ 越大,集合 X 的不确定程度越大。近似精度和粗糙度只考虑了系统的不确定性,不能完全反映粗集的不确定性,基于知识含量的近似精度和粗糙度是处理粗集概念不确定性的理想指标。

知识 R 含量的测度：$I(R) = 1 - \frac{1}{|U|^2}\sum_{i=1}^{n}|X_i|^2$ (2.2)

集合 X 基于知识含量 R 的粗糙度：$\text{Rough}_R(X) = \rho_R(X)(1 - I(R))$

集合 X 基于知识含量 R 的近似精度：$D_R(X) = 1 - \text{Rough}_R(X)$

(2.3)

公式(2.2)中 $|X_i|$ 表示集合 X_i 的基数，n 为知识库 $(U,R) = \{X_1, X_2, \cdots, X_n\}$ 中知识的数量，$i = 1, 2, \cdots, n$，将公式(2.2)代入公式(2.3)，得到

$$D_R(X) = I(R) + \alpha_R(X) - I(R)\alpha_R(X)$$

知识库划分越细、越均匀，知识 R 含量的测度 $I(R)$ 的值就越大。集合 X 的基于知识含量 R 的近似精度 $D_R(X)$ 和基于知识含量 R 的粗糙度 $\text{Rough}_R(X)$ 均是一个区间[0,1]上的实数，它们分别定义了集合 X 的可定义程度和知识的不完全程度。$D_R(X)$ 越大，集合 X 的确定度越大；与近似精度 $D_R(X)$ 恰恰相反，$\text{Rough}_R(X)$ 越大，集合 X 的不确定程度越大。

2.2 空间数据的不确定性模型

空间数据的随机不确定性表现为地理信息的位置不确定性，模糊不确定性表现为地理信息属性和语义的不确定性，具有位置不确定性的地理目标为随机目标，具有模糊不确定性的地理目标为模糊空间目标。

在随机目标的表达方面，其基本假设前提是点位误差服从正态或圆正态分布。基于此，Chrisman(1982)、刘文宝(1998)、蓝悦明(2003)等分别提出了"ε-带"、"G-带"、"C 环"等随机线、面目标表达方法。在模糊空间目标的表达方面，有关模糊目标的形式化描述成为研究焦点，D. Attman 等人(1994)分别基于点集拓扑、模糊集的 α 截集、有限精度的空间格网划分、模糊胞腔和模糊胞腔复形，提出了模糊面状目标的形式化表达方法，何建华(2005)基于模糊贴近度构建了能集成表达模糊点、模糊线和模糊面的位置不确定性和主题

属性模糊性（没有考虑语义模糊性）的形式化描述方法。

以上关于空间目标的不确定性表达方法中，一些附加信息或先验信息如模糊集理论中的成员隶属函数、证据理论中的基本概率赋值和统计学中的概率分布等通常是难以得到和确定的。一些学者又研究运用粗集理论对空间数据的不确定性进行分析建模。例如，王树良等人基于粗集的思想方法定义了粗实体、粗关系和粗算子。邓敏（2006）运用粗集理论中的一些基本概念和方法系统地描述和表达了空间目标位置数据、属性数据以及空间关系数据的不确定性。

随着考虑不确定性的类型以及分析处理时的侧重点不同，可以利用不同的数学工具来分析和表达。而且，这些空间数据不确定性模型存在一定的互通性和相似性。例如，模糊集中的支撑集和核集（core）与粗集中的上近似和下近似在一定程度上具有等价性，误差带（环）可以作为粗集的边界区域等。

2.2.1　随机目标位置误差模型

点位精度是测绘学科研究的重要问题，图 2.1 所示的"ε-带"模型一直是研究线的不确定性问题的基础。最初的"ε-带"模型是针对一条线的简单、对称的缓冲区，没有考虑组成的点与线之间存在的关系。逐渐地，"ε-带"的形状和概率成为研究热点。刘大杰、史文中、童小华等人利用"ε-带"理论，对 GIS 中的点、直线、圆曲线、一般曲线、面的不确定性进行了研究，建立了误差模型，导出了误差带数学方程，分析了误差带形状特征并加以可视化表达。同时，很多学者运用模糊集合理论、概率论等对此进行了修正和发展。进而产生了"E-带"、"S-带"、"G-带"模型。这些误差带均是与一定置信水平相关的置信区间，为了得到不随概率变化的误差带，范爱民等人（2001）、李大军等人（2003）利用信息论建立了一种称为误差熵的新型误差带——"H-带"，即线元的真实位置位于该确定的区域内。

本小节将对随机点、线、面目标的位置不确定性模型中最基础的模型进行论述和分析。

1. 点元的位置不确定性模型

点目标是 GIS 中最基本的描述对象之一。点位精度是测绘学科

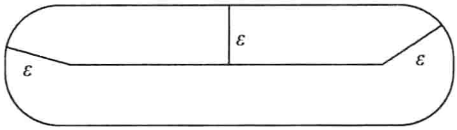


图 2.1　“ε-带”

研究的重要问题，平面点坐标的不确定性一般用方差或标准差（中误差）来表示，而用点位中误差和误差椭圆（或中误差曲线）、置信椭圆、可靠度椭圆表示平面点位的精度或不确定性。坐标中误差和点位中误差及误差椭圆、置信椭圆、可靠度椭圆完整地描述了二维点的误差在各个方向上的分布状况。

设二维随机点 Z 服从标准正态分布，点 $Z(x,y)$ 在 x 方向方差为 σ_x^2、标准差为 σ_x，y 方向方差为 σ_y^2、标准差为 σ_y，两者的协方差为 σ_{xy}，且随机点元在坐标轴 x、y 方向的相关系数为 ρ_{xy}，点位中误差为 $\sigma_Z = \sqrt{\sigma_x^2 + \sigma_y^2}$。

误差椭圆是用于描述点位误差最常用的方法。若随机点 Z 服从标准正态分布，则其概率密度分布函数可以确定为：

$$f(x,y) = \frac{\exp \dfrac{-\left[\dfrac{(x-\mu_x)^2}{\sigma_x^2} + \dfrac{(y-\mu_y)^2}{\sigma_y^2} - \dfrac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y}\right]}{2(1-\rho_{xy}^2)}}{2\pi \sigma_x \sigma_y \sqrt{1-\rho_{xy}^2}}$$

且相应标准误差椭圆的长半轴 E、短半轴 F 以及主轴方向角 θ 可以确定为：

$$E = \frac{\sqrt{\sigma_x^2 + \sigma_y^2 + [(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2]}}{2}$$

$$F = \frac{\sqrt{\sigma_x^2 + \sigma_y^2 - [(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2]}}{2}$$

$$\theta = \frac{1}{2}\arctan\left(\frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}\right)$$

这样，点目标落入标准误差椭圆的概率值 P 即为该标准误差椭

圆对应的二维误差曲面包绕空间的体积值，如图 2.2 所示。

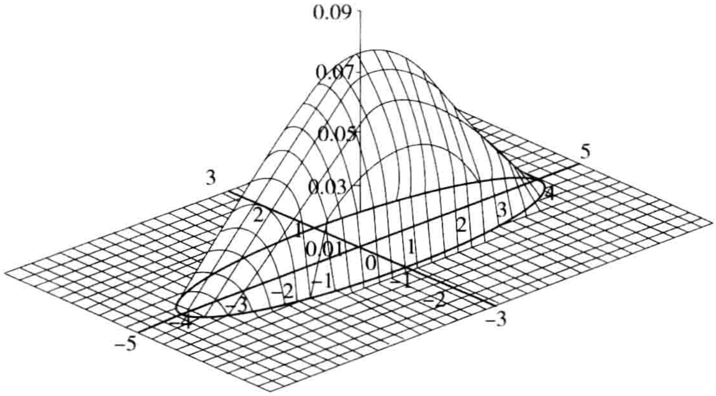


图 2.2　标准误差椭圆对应的二维误差曲面(史文中等,2005)

概率值随误差椭圆大小的增大而增大。例如，当误差椭圆扩大到两倍标准误差椭圆时，概率值为 0.893；当误差椭圆扩大到三倍标准误差椭圆时，点元落入相应的误差椭圆内的概率值为 0.989。

这种确定的概率值广泛地应用于描述点元的位置不确定性。另外，这类方法也进一步地延伸至随机线元和多边形的位置误差的评估。

2. 线元位置不确定性模型

(1)“ε_σ- 带”模型

假设如图 2.3 所示的直线的两个端点坐标分别为 $P_1(x_1, y_1)$ 和 $P_2(x_2, y_2)$，它们的方差 - 协方差矩阵为：

$$D_{12} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1y_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} \\ \sigma_{x_1y_1} & \sigma_{y_1}^2 & \sigma_{y_1x_2} & \sigma_{y_1y_2} \\ \sigma_{x_1x_2} & \sigma_{x_2y_1} & \sigma_{x_2}^2 & \sigma_{x_2y_2} \\ \sigma_{x_1y_2} & \sigma_{y_1y_2} & \sigma_{x_2y_2} & \sigma_{y_2}^2 \end{bmatrix}$$

$\sigma_{s_i}^2$、$\sigma_{s_is_j}$ 分别为量测距离的方差和协方差值。

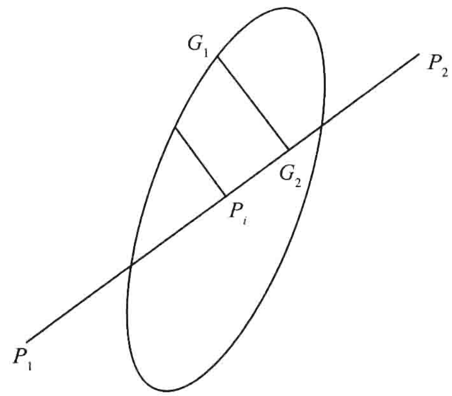


图 2.3　直线上任意一点的误差

对于线段P_1P_2上任意点P_i的坐标可通过点P_1由下式得到

$$\begin{cases} x_i = x_1 + S_{1i}\cos\alpha_{1i} = \left(1 - \dfrac{S_{1i}}{S_{12}}\right)x_1 + \dfrac{S_{1i}}{S_{12}}x_2 \\ y_i = y_1 + S_{1i}\sin\alpha_{1i} = \left(1 - \dfrac{S_{1i}}{S_{12}}\right)y_1 + \dfrac{S_{1i}}{S_{12}}y_2 \end{cases} \tag{2.4}$$

式(2.4)中S_{ij}、α_{ij}分别表示线段P_1P_2上任意两点P_i、P_j间的距离和方位角。

如果不考虑S_{12}和S_{1i}的误差，则由式(2.4)微分得，

$$\begin{cases} \mathrm{d}x_i = (1 - r_1)\mathrm{d}x_1 + r_1\mathrm{d}x_2 \\ \mathrm{d}y_i = (1 - r_1)\mathrm{d}y_1 + r_1\mathrm{d}y_2 \end{cases} \tag{2.5}$$

式中，$r_1 = \dfrac{S_{1i}}{S_{12}}$。

由式(2.5)应用协方差传播律可得x_i、y_i的方差$\sigma_{x_i}^2$、$\sigma_{y_i}^2$和协方差$\sigma_{x_iy_i}$，

$$\sigma_{x_i}^2 = (1 - r)^2\sigma_{x_1}^2 + r^2\sigma_{x_2}^2 + 2r(1 - r)\sigma_{x_1x_2}$$

$$\sigma_{y_i}^2 = (1 - r)^2\sigma_{y_1}^2 + r^2\sigma_{y_2}^2 + 2r(1 - r)\sigma_{y_1y_2}$$

$$\sigma_{x_iy_i} = (1 - r)^2\sigma_{x_1y_1} + r^2\sigma_{x_2y_2} + 2r(1 - r)(\sigma_{x_1y_2} + \sigma_{x_2y_1})$$

在求得平面点$P_i(x_i,y_i)$的方差($\sigma_{x_i}^2,\sigma_{y_i}^2$)和协方差$\sigma_{x_iy_i}$后,可以求得该点在任意方向的方差,并得到该点的中误差曲线和误差椭圆。Gaspry 等人(1992)、Shi (1994)以该点在垂直于线段的方向(X'方向)上的中误差$\sigma_{x_i'}$作为"ε-带"误差带宽,作为反映线段上各点的误差分布状况的不确定性模型,也就是"ε_σ-带"模型。

为了求定$\sigma_{x'}$值,先将原坐标系xOy进行旋转,使旋转后的坐标系$x'Oy'$的y'轴与线段方向一致,如图 2.4 所示,旋转角 $\theta = \arctan\dfrac{x_2 - x_1}{y_2 - y_1}$,则$P_i$在$x'Oy'$坐标系中的坐标为:

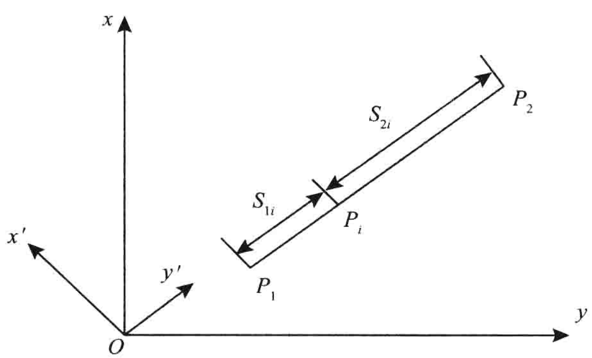


图 2.4　直线及直线上任意一点 i

$$[x'_i, y'_i] = [x_i, y_i]\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

由此可得,(x'_i, y'_i)的方差$\sigma_{x'_i}^2$、$\sigma_{y'_i}^2$和协方差$\sigma_{x'_iy'_i}$为:

$$\sigma_{x'_i}^2 = \cos^2\theta \cdot \sigma_{x_i}^2 - \sin2\theta \cdot \sigma_{x_iy_i} + \sin^2\theta \cdot \sigma_{y_i}^2$$

$$\sigma_{y'_i}^2 = \sin^2\theta \cdot \sigma_{x_i}^2 - \sin2\theta \cdot \sigma_{x_iy_i} + \cos^2\theta \cdot \sigma_{y_i}^2$$

$$\sigma_{x'_iy'_i} = \sin\theta\cos\theta \cdot \sigma_{x_i}^2 + \cos2\theta \cdot \sigma_{x_iy_i} - \sin\theta\cos\theta \cdot \sigma_{y_i}^2$$

垂直于线段方向的中误差σ'_{x_i}即为"ε_σ-带"模型,这也是 Caspry 等人(1992)、Shi (1994)所取的"ε-带"宽值。

(2)"ε_m-带"模型

如图 2.3 所示,P_i点处误差椭圆至线段距离最大的点在 G_1 误差

点,也就是说,点在与线段相垂直的方向(x' 方向)上的中误差不一定是误差曲线(或误差椭圆)上至线段的最大距离。因此,取图 2.3 中点 G_1 至线段的距离 G_1G_2 为误差带宽值建立误差带为"ε_m-带"。

在$x'Oy'$ 中,P_i 在方位角为 φ 的任意方向上的方差为:

$$\sigma_\varphi^2 = \cos^2\varphi \cdot \sigma_x^2 + \sin^2\varphi \cdot \sigma_y^2 + \sin2\varphi \cdot \sigma_{xy}$$

误差椭圆上的点,当它到直线的垂直距离最大时,其$\sigma_\varphi \cos\varphi$ 达到极大值,也就是

$$\frac{\mathrm{d}(\sigma_\varphi^2 \cos^2\varphi)}{\mathrm{d}\varphi} = \frac{\mathrm{d}(\cos^4\varphi \cdot \sigma_x^2 + \sin^2\varphi \cos^2\varphi \cdot \sigma_y^2 + \sin2\varphi\, \sigma_{xy} \cos^2\varphi)}{\mathrm{d}\varphi} = 0 \tag{2.6}$$

对式(2.6)进行整理、化简、求解,可以求出误差椭圆上到直线上垂直距离最大的点的 φ 值,然后求出对应于此点的ε_m 值为:

$$\varepsilon_m = \sigma_\varphi \cdot \cos\varphi$$

ε_m 沿直线线元所得的轨迹即组成包绕该线元的最大误差区域。

(3)以点位误差表示的误差带

"ε_σ-带"和"ε_m-带"是按误差椭圆簇形成的误差带,理论上是严密的,能够反映出线元的误差状况,可视化效果也很好,但是,其公式非常繁琐,计算复杂,于是蓝悦明等人(2004)将其进行改进,用点位误差来表示线元的误差带,该方法保存了原有误差带的优点,又简化了计算公式。

一般情况下,线元P_1、P_2 两点构成的直线上任意点P_i 的点位精度可表达为:

$$\sigma_i^2 = \sigma_{x_i}^2 + \sigma_{y_i}^2 = (1-r)^2\sigma_1^2 + r^2\sigma_2^2 + 2r(1-r)(\sigma_{x_1x_2} + \sigma_{y_1y_2}) \tag{2.7}$$

式中,$\sigma_1^2 = \sigma_{x_1}^2 + \sigma_{y_1}^2$ 为端点P_1 的点位精度;$\sigma_2^2 = \sigma_{x_2}^2 + \sigma_{y_2}^2$ 为端点P_2 的点位精度。以式(2.7)为基础,以 r 为参变量计算线元上各处的σ_i,然后构成误差带。点位误差具有与误差椭圆相同的表示误差大小的作用,同一点处的点位方差σ_i 与误差椭圆的长半轴A_r、短半轴B_r 有直接的关系,即$\sigma_i^2 = A_r^2 + B_r^2$。

当$\sigma_{x_1x_2} = \sigma_{y_1y_2} = 0$ 时,式(2.7)简化为:

$$\sigma_i^2 = (1-r)^2\sigma_1^2 + r^2\sigma_2^2$$

3. 面元位置误差模型

面的不确定性是组成它的各线段的误差带的集合，因此，这里用线元的“ε_σ- 带”模型和“ε_m- 带”模型来描述面的误差带(环)边界线方程。图 2.5 是由折线所组成的封闭多边形。对其中一段线元 $P_i P_{i+1}$ 上任意一点，取垂直于线段方向的中误差作为其不确定性的带宽值，即ε_σ 值。设(x_{ir}, y_{ir})是线元$P_i P_{i+1}$ 上比值为r_i 处对应点的坐标，σ_{ir} 为点(x_{ir}, y_{ir})处的ε_σ 值，则

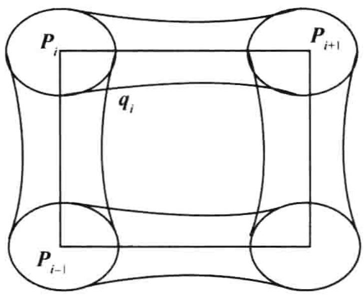


图 2.5　折线组成的封闭多边形

$$\sigma_{ir} = \sqrt{(1 - r_i)^2 \sigma_{x_i}^2 + r_i^2 \sigma_{x_{i+1}}^2 + 2 r_i(1 - r_i) \sigma_{x_i x_{i+1}}}$$

设线元$P_i P_{i+1}$ 的方位角为$\alpha_{i,i+1}$，则线元$P_i P_{i+1}$ 的“ε_σ- 带”的外边界线上任意一点的参数方程为：

$$\begin{cases} x = x_{ir} + \sigma_{ir} \sin \alpha_{i,i+1} \\ y = y_{ir} - \sigma_{ir} \cos \alpha_{i,i+1} \end{cases} \tag{2.8}$$

又

$$\begin{cases} x_{ir} = (1 - r_i) x_i + r_i x_{i+1} \\ y_{ir} = (1 - r_i) y_i + r_i y_{i+1} \end{cases}$$

因此，面元误差带的外边界线包括线元误差带和端点P_i、P_{i+1} 误差椭圆的一部分，如图 2.5 所示。

对内边界线，设线元$P_i P_{i+1}$ 与线元$P_{i-1} P_i$ 误差带的交点为q_i，则误差带上任意一点的参数方程为

$$\begin{cases} x = x_{ir} - \sigma_{ir}\sin\alpha_{i,i+1} \\ y = y_{ir} + \sigma_{ir}\cos\alpha_{i,i+1} \\ x_{q_{i-1}} \leqslant x \leqslant x_{q_i} \\ y_{q_{i-1}} \leqslant y \leqslant y_{q_i} \end{cases}$$

交点$q_i(x_{q_i}, y_{q_i})$可以由式(2.8)求得，

$$\begin{cases} x_{q_i} = x_{ir} + \sigma_{ir}\sin\alpha_{i,i+1} = x_{(i-1)_r} + \sigma_{(i-1)_r}\sin\alpha_{i-1,i} \\ y_{q_i} = y_{ir} - \sigma_{ir}\cos\alpha_{i,i+1} = y_{(i-1)_r} - \sigma_{(i-1)_r}\cos\alpha_{i-1,i} \end{cases} \tag{2.9}$$

同理，也可得到线元的ε_m模型下的误差环的边界线方程。图2.6是面元的不确定性误差带形状。

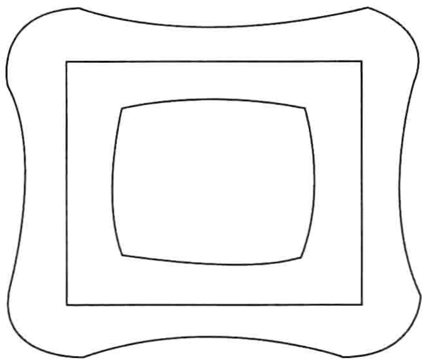


图 2.6 面元的不确定性误差带模型

对面元的外边界仍由线元对应的误差带和端点所对应的部分误差椭圆所构成。对线元$P_i P_{i+1}$上任意一点取其曲线上至线段的最大距离作为线元不确定性的带宽值，即ε_m值，

$$\varepsilon_m = \max_{0\leqslant\varphi\leqslant2\pi}(\sigma_\varphi\cos\varphi) = \sigma_{\theta_r}\cos\theta_r \tag{2.10}$$

式(2.10)中，θ_r为线元$P_i P_{i+1}$上任意一点处误差椭圆上到线元$P_i P_{i+1}$最大距离的点与线元的垂线段的方位角。这样，其外边界线误差带上任意一点的参数方程为：

$$\begin{cases} x = x_{ir} + \sigma_{\theta_r}\cos\theta_r \\ y = y_{ir} + \sigma_{\theta_r}\sin\theta_r \end{cases} \tag{2.11}$$

式(2.11)中(x_{ir}, y_{ir})同式(2.9)。

对内边界线,误差带上任意一点的参数方程为:

$$\begin{cases} x = x_{ir} - \sigma_{\theta_r}\cos\theta_r \\ y = y_{ir} - \sigma_{\theta_r}\sin\theta_r \\ \quad x_{q_{i-1}} \leqslant x \leqslant x_{qi} \\ \quad y_{q_{i-1}} \leqslant y \leqslant y_{qi} \end{cases}$$

类似地,线元$P_i P_{i+1}$与线元$P_{i-1} P_i$误差带的交点q_i,由下式求得

$$\begin{cases} x_{q_i} = x_{ir} - \sigma_{\theta_r}\cos\theta_r = x_{(i-1)_r} - \sigma_{(i-1)\theta_r}\cos(i-1)\theta_r \\ y_{q_i} = y_{ir} - \sigma_{\theta_r}\sin\theta_r = y_{(i-1)_r} - \sigma_{(i-1)\theta_r}\sin(i-1)\theta_r \end{cases} \tag{2.12}$$

式(2.12)中,$(i-1)\theta_r$表示线元$P_{i-1} P_i$上任意一点最大误差所对应的方位角。

面元误差带的面积值也可以作为衡量其不确定性的一个重要指标。这里以图2.5所示的矩形多边形为例,并假设各端点等精度,来计算图2.6所示的面元误差带的面积。

记各线段误差带面积为A_i,由面元的内边界线和误差圆所围成的近似三角形面积记为Δ_i,则该面元的误差带面积为:

$$\text{Area} = 4(A_i - \pi R^2 - \Delta_i)$$

其中

$$R = \sigma_{x_i}$$

$$A_i = 2L\int_0^1 \sigma_r \mathrm{d}r + \pi R^2$$

$$\Delta_i = L\int_0^{qi} \sigma_{(i)r}\mathrm{d}r + L\int_{qi}^{\frac{R}{L}} \sigma_{(i-1)r}\mathrm{d}r - \frac{1}{4}\pi R^2$$

式中σ_r可取ε_σ或ε_m。

这里面元误差带的面积Area并不是该面元的面积的中误差。

4. 一般曲线的不确定性模型

(1) 一般曲线上任意点的位置不确定性

设一般曲线的函数方程或拟合方程为:

$$f(x, y, \alpha, \beta) = 0 \tag{2.13}$$

其中(x, y)是曲线上的一个点,$\alpha = [\alpha_1\ \alpha_2]^{\mathrm{T}}$和$\beta = [\beta_1\ \beta_2]^{\mathrm{T}}$为曲线

的参数矢量。对于曲线上两点k_1、k_2及它们之间的任意一点k_i，有

$$\begin{cases} f(x_1, y_1, \alpha, \beta) = 0 \\ f(x_2, y_2, \alpha, \beta) = 0 \\ f(x_i, y_i, \alpha, \beta) = 0 \end{cases} \tag{2.14}$$

设点k_1至k_i的观测距离为S_{1i}，点k_2至k_i的观测距离为S_{2i}，则

$$\begin{cases} S_{1i}^2 = (x_i - x_1)^2 + (y_i - y_1)^2 \\ S_{2i}^2 = (x_i - x_2)^2 + (y_i - y_2)^2 \end{cases} \tag{2.15}$$

将式(2.14)、式(2.15)进行微分，得

$$\begin{cases} A_{\alpha_1}\delta\alpha - A_{\beta_1}\delta\beta + a_{x_1}\mathrm{d}x_1 + a_{y_1}\mathrm{d}y_1 = 0 \\ A_{\alpha_2}\delta\alpha - A_{\beta_2}\delta\beta + a_{x_2}\mathrm{d}x_2 + a_{y_2}\mathrm{d}y_2 = 0 \\ A_{\alpha_i}\delta\alpha - A_{\beta_i}\delta\beta + a_{x_i}\mathrm{d}x_i + a_{y_i}\mathrm{d}y_i = 0 \\ S_{1i}\mathrm{d}S_{1i} = \Delta x_{1i}(\mathrm{d}x_i - \mathrm{d}x_1) + \Delta y_{1i}(\mathrm{d}y_i - \mathrm{d}y_1) \end{cases} \tag{2.16}$$

式中$\delta\alpha = [\delta\alpha_1\ \delta\alpha_2 \cdots]^{\mathrm{T}}$，$\delta\beta = [\delta\beta_1\ \delta\beta_2]^{\mathrm{T}}$。

将式(2.16)进行整理，可得

$$\begin{bmatrix} \mathrm{d}x_i' \\ \mathrm{d}y_i' \end{bmatrix} = \begin{bmatrix} A_x & A_{xs} & 0 \\ A_y & A_{ys} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{d}\eta \\ \mathrm{d}S_{1i} \\ \mathrm{d}S_{2i} \end{bmatrix} = A\mathrm{d}\xi \tag{2.17}$$

式(2.17)中，$\mathrm{d}\eta = [\mathrm{d}\bar{X}\quad \delta\alpha]^{\mathrm{T}}$

$$A = \begin{bmatrix} A_x & A_{xs} & 0 \\ A_y & A_{ys} & 0 \end{bmatrix} = \begin{bmatrix} \Delta x_{1i} & \Delta y_{1i} \\ a_{x(i)} & a_{y(i)} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_{i1} & S_{1i} & 0 \\ \alpha_{i2} & 0 & 0 \end{bmatrix}$$

$$\alpha_{i1} = [\Delta x_{1i}, \Delta y_{1i}, 0, 0, 0]$$

$$\alpha_{i2} = \begin{bmatrix} 0 \\ -A_{\alpha_i} \end{bmatrix} + A_{\beta_i} \begin{bmatrix} A_{\beta_1} \\ A_{\beta_2} \end{bmatrix}^{-1} \begin{bmatrix} a_{x_1} & a_{y_1} & 0 & 0 & A_{\alpha_1} \\ 0 & 0 & a_{x_2} & a_{y_2} & A_{\alpha_2} \end{bmatrix}$$

$$\mathrm{d}\xi = [\mathrm{d}x_1\ \mathrm{d}y_1\ \mathrm{d}x_2\ \mathrm{d}y_2\ \delta\alpha\ \mathrm{d}S_{1i}\ \mathrm{d}S_{2i}]^{\mathrm{T}}$$

若考虑点k_2至k_i的观测距离S_{2i}，同理可得出关于(x_i, y_i)的方程。

$$\begin{bmatrix} \mathrm{d}x''_i \\ \mathrm{d}y''_i \end{bmatrix} = \begin{bmatrix} B_x & 0 & B_{xs} \\ B_y & 0 & B_{ys} \end{bmatrix} \begin{bmatrix} \mathrm{d}\eta \\ \mathrm{d}S_{1i} \\ \mathrm{d}S_{2i} \end{bmatrix} = B\mathrm{d}\xi \tag{2.18}$$

式(2.18)中

$$B = \begin{bmatrix} B_x & 0 & B_{xs} \\ B_y & 0 & B_{ys} \end{bmatrix} = \begin{bmatrix} \Delta x_{2i} & \Delta y_{2i} \\ a_{x(i)} & a_{y(i)} \end{bmatrix}^{-1} \begin{bmatrix} \beta_{i1} & 0 & S_{2i} \\ \beta_{i2} & 0 & 0 \end{bmatrix}$$

$$\beta_{i1} = [\Delta x_{2i}, \Delta y_{2i}, 0, 0, 0]$$

$$\beta_{i2} = \begin{bmatrix} 0 \\ -A_{\alpha_i} \end{bmatrix} + A_{\beta_i} \begin{bmatrix} A_{\beta_1} \\ A_{\beta_2} \end{bmatrix}^{-1} \begin{bmatrix} a_{x_1} & a_{y_1} & 0 & 0 & A_{\alpha_1} \\ 0 & 0 & a_{x_2} & a_{y_2} & A_{\alpha_2} \end{bmatrix}$$

(2)一般曲线位置不确定性的“ε_σ-带”模型

这里考虑将曲线上点k_1、k_2之间的任意一点k_i的不确定性用其坐标的方差表示。对微分关系式(2.17)、式(2.18)用权$\frac{1}{\hat{S}_{1i}}$和$\frac{1}{\hat{S}_{2i}}$进行加权,并令$r_1 = \frac{\hat{S}_{1i}}{\hat{S}_{12}}$,则

$$\begin{bmatrix} \mathrm{d}x_i \\ \mathrm{d}y_i \end{bmatrix} = (1 - r_1) \begin{bmatrix} \mathrm{d}x'_i \\ \mathrm{d}y'_i \end{bmatrix} + r_1 \begin{bmatrix} \mathrm{d}x''_i \\ \mathrm{d}y''_i \end{bmatrix} = (A - r_1 A + r_1 B) \cdot \mathrm{d}\xi \tag{2.19}$$

式(2.19)中,如果点k_1、k_2为一条直线,则有$r_1 = \frac{S_{1i}}{S_{12}}$,不考虑距离量测误差,式(2.19)便化为:

$$\begin{cases} \mathrm{d}x_i = \frac{S_{2i}}{S_{12}} \mathrm{d}x'_i + \frac{S_{1i}}{S_{12}} \mathrm{d}x''_i = (1 - r_1)\,\mathrm{d}x_1 + r_1\,\mathrm{d}x_2 \\ \mathrm{d}y_i = \frac{S_{2i}}{S_{12}} \mathrm{d}y'_i + \frac{S_{1i}}{S_{12}} \mathrm{d}y''_i = (1 - r_1)\,\mathrm{d}y_1 + r_1\,\mathrm{d}y_2 \end{cases}$$

上式即为 GIS 线元不确定性的一般模型。

根据协方差传播律可得到

$$D_i = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_i y_i} \\ \sigma_{x_i y_i} & \sigma_{y_i}^2 \end{bmatrix} = (A - r_1 A + r_1 B) \cdot D\xi \cdot (A - r_1 A + r_1 B)^{\mathrm{T}}$$

式中，$\sigma_{x_i}^2$ 和$\sigma_{y_i}^2$ 是k_i 点的坐标方差，D_ξ 表示 ξ 的方差阵。设曲线上k_i 点法线方向的方位角为θ，以其法线方向的中误差σ_θ 表示误差带宽，由于

$$u_\theta = \cos\theta \cdot x_i + \sin\theta \cdot y_i$$

则

$$\sigma_\theta^2 = \cos^2(\theta) \cdot \sigma_{x_i}^2 + \sin^2(\theta) \cdot \sigma_{y(i)}^2 + \sin2\theta \cdot \sigma_{x_iy_i}$$

曲线上各点σ_θ 的轨迹就构成了曲线的不确定性范围，称之为“ε_σ-带”。

（3）最大方向的误差带“ε_m-带”模型

曲线的法线方向的中误差σ_θ 是$k_i(x_i, y_i)$ 点误差椭圆与曲线的法线交点G_θ 至k_i 的距离，如图2.7所示，k_i 点误差椭圆上至曲线k_1k_2 的距离最大点在G_φ 点处，因此，由G_φ 向曲线k_1k_2 作法线与其交于点G_λ，取$|G_\varphi G_\lambda|$ 来表征曲线的不确定性。

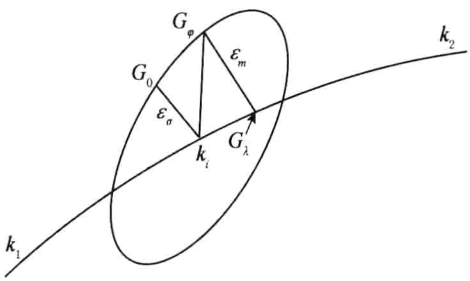


图2.7 一般曲线最大方向误差

设k_iG_φ 的方位角为 φ，有

$$\sigma_\varphi^2 = \cos^2\varphi \cdot \sigma_{x_i}^2 + \sin^2\varphi \cdot \sigma_{y_i}^2 + \sin2\varphi \cdot \sigma_{x_iy_i} \tag{2.20}$$

G_φ 的坐标为：

$$\begin{cases} x_\varphi = x_i + \sigma_\varphi\cos\varphi \\ y_\varphi = y_i + \sigma_\varphi\sin\varphi \end{cases} \tag{2.21}$$

过点$G_\lambda(x_\lambda, y_\lambda)$ 的曲线k_1k_2 的法线方程为：

$$\frac{x - x_\lambda}{f'_{x_\lambda}} - \frac{y - y_\lambda}{f'_{y_\lambda}} = 0 \tag{2.22}$$

式中

$$f'_{x_\lambda} = \left.\frac{\partial f}{\partial x}\right|_{(x_\lambda, y_\lambda)}$$

$$f'_{y_\lambda} = \left.\frac{\partial f}{\partial y}\right|_{(x_\lambda, y_\lambda)}$$

因为点$G_\varphi(x_\varphi, y_\varphi)$在法线上，故有

$$(x_i - x_\lambda + \sigma_\varphi \cos\varphi) f'_{y_\lambda} - (y_i - y_\lambda + \sigma_\varphi \sin\varphi) f'_{x_\lambda} = 0 \tag{2.23}$$

设$G_\varphi\ G_\lambda$的距离为$S_{\varphi\lambda}$，有

$$S^2_{\varphi\lambda} = (x_\varphi - x_\lambda)^2 + (y_\varphi - y_\lambda)^2 = (x_i - x_\lambda + \sigma_\varphi \cos\varphi)^2 + (y_i - y_\lambda + \sigma_\varphi \sin\varphi)^2 \tag{2.24}$$

为了求得$S_{\varphi\lambda}$的最大值及相应的φ，将φ、σ_φ和x_λ，y_λ作为参数，而将式(2.13)、式(2.20)、式(2.23)作为参数应满足的条件，分别对φ、σ_φ、x_λ、y_λ取近似值φ^0、σ^0_φ和x^0_λ、y^0_λ，将式(2.13)、式(2.20)和式(2.23)用泰勒级数展开，并仅取一次项，构成如下的条件式

$$\begin{cases} \alpha_{\varphi 1}\delta\varphi + \alpha_{\sigma 1}\delta\sigma_\varphi + w_1 = 0 \\ \alpha_{\varphi 2}\delta\varphi + \alpha_{\sigma 2}\delta\sigma_\varphi + \alpha_{x_\lambda 2}\delta x_\lambda + \alpha_{y_\lambda 2}\delta y_\lambda + w_2 = 0 \\ \alpha_{x_\lambda 3}\delta x_\lambda + \alpha_{y_\lambda 3}\delta y_\lambda + w_3 = 0 \end{cases} \tag{2.25}$$

式中

$$\alpha_{\varphi 1} = \sigma^2_{x_i}\sin 2\varphi^0 - \sigma^2_{y_i}\sin 2\varphi^0 - 2\sigma_{x_i y_i}\cos 2\varphi^0$$

$$\alpha_{\sigma 1} = 2\sigma^0_\varphi$$

$$w_1 = (\sigma^0_\varphi)^2 - (\sigma^2_{x_i}\cos^2\varphi^0 + \sigma^2_{y_i}\sin^2\varphi^0 + \sigma_{x_i y_i}\sin 2\varphi^0)$$

$$\alpha_{\varphi 2} = -\sin\varphi^0\,\sigma^0_\varphi\,(f'_{y_\lambda})^0 - \cos\varphi^0\,\sigma^0_\varphi\,(f'_{x_\lambda})^0$$

$$\alpha_{\sigma 2} = \cos\varphi^0\,(f'_{y_\lambda})^0 + \sin\varphi^0\,(f'_{x_\lambda})^0$$

$$\alpha_{x_\lambda 2} = -(f'_{y\lambda})^0 + (x_i - x^0_\lambda + \sigma^0_\varphi\cos\varphi^0)(f''_{y\lambda})^0 - (y_i - y^0_\lambda + \sigma^0_\varphi\sin\varphi^0)(f''_{x\lambda})^0$$

$$\alpha_{y_\lambda 2} = (f'_{x\lambda})^0 + (x_i - x_\lambda^0 + \sigma_\varphi^0 \cos\varphi^0)(f''_{y\lambda})^0 - (y_i - y_\lambda^0 + \sigma_\varphi^0 \sin\varphi^0)(f''_{x\lambda})^0$$

$$w_2 = (x_i - x_\lambda^0 + \sigma_\varphi^0 \cos\varphi^0)(f''_{y\lambda})^0 - (y_i - y_\lambda^0 + \sigma_\varphi^0 \sin\varphi^0)(f''_{x\lambda})^0$$

$$\alpha_{x_\lambda 3} = (f'_{x\lambda})^0$$

$$\alpha_{y_\lambda 3} = (f'_{y\lambda})^0$$

$$w_3 = f^0$$

令

$$\begin{cases} u_1 = x_i - x_\lambda + \sigma_\varphi \cos\varphi \\ u_2 = y_i - y_\lambda + \sigma_\varphi \sin\varphi \end{cases} \tag{2.26}$$

展开,得

$$\begin{cases} u_1 = u_1^0 + \cos\varphi^0\,\delta\sigma_\varphi - \sigma_\varphi^0 \sin\varphi^0 \delta\varphi - \delta x_\lambda \\ u_2 = u_2^0 + \sin\varphi^0\,\delta\sigma_\varphi - \sigma_\varphi^0 \cos\varphi^0 \delta\varphi - \delta y_\lambda \end{cases} \tag{2.27}$$

按求条件极值的方法构成目标函数

$$\Omega = u_1^2 + u_2^2 + 2k_1(\alpha_{\varphi 1}\delta\varphi + \alpha_{\sigma 1}\delta\sigma_\varphi + w_1) + 2k_2(\alpha_{\varphi 2}\delta\varphi + \alpha_{\sigma 2}\delta\sigma_\varphi + \alpha_{x_\lambda 2}\delta x_\lambda + \alpha_{y_\lambda 2}\delta y_\lambda + w_2) + 2k_3(\alpha_{x_\lambda 3}\delta x_\lambda + \alpha_{y_\lambda 3}\delta y_\lambda + w_3) \tag{2.28}$$

由

$$\frac{\partial\Omega}{\partial\delta\varphi} = -2u_1\sigma_\varphi^0\sin\varphi^0 + 2u_2\sigma_\varphi^0\cos\varphi^0 + 2k_1\alpha_{\varphi 1} + 2k_2\alpha_{\varphi 2} = 0$$

$$\frac{\partial\Omega}{\partial\delta\sigma_\varphi} = 2u_1\cos\varphi^0 + 2u_2\sin\varphi^0 + 2k_1\alpha_{\sigma 1} + 2k_2\alpha_{\sigma 2} = 0$$

$$\frac{\partial\Omega}{\partial\delta x_\lambda} = -2u_1 + 2k_2\alpha_{x_\lambda 2} + 2k_3\alpha_{x_\lambda 3} = 0$$

$$\frac{\partial\Omega}{\partial\delta y_\lambda} = -2u_2 + 2k_2\alpha_{y_\lambda 2} + 2k_3\alpha_{y_\lambda 3} = 0$$

整理可得方程组

$$\begin{cases} (\sigma_\varphi^0)^2\delta\varphi + \sigma_\varphi^0\sin\varphi^0\delta x_\lambda - \sigma_\varphi^0\cos\varphi^0\delta y_\lambda + \alpha_{\varphi 1}k_1 + \alpha_{\varphi 2}k_2 + \\ \quad (-u_1^0\sigma_\varphi^0\sin\varphi^0 + u_2^0\sigma_\varphi^0\cos\varphi^0) = 0 \\ \delta\sigma_\varphi - \cos\varphi^0\delta x_\lambda - \sin\varphi^0\delta y_\lambda + \alpha_{\sigma 1}k_1 + \alpha_{\sigma 2}k_2 + \\ \quad (u_1^0\cos\varphi^0 + u_2^0\sin\varphi^0) = 0 \\ \sin\varphi^0\sigma_\varphi^0\delta\varphi - \cos\varphi^0\delta\sigma_\varphi + \delta x_\lambda + \alpha_{x_\lambda 2}k_2 + \alpha_{x_\lambda 3}k_3 - u_1^0 = 0 \\ -\cos\varphi^0\sigma_\varphi^0\delta\varphi - \sin\varphi^0\delta\sigma_\varphi + \delta y_\lambda + \alpha_{y_\lambda 2}k_2 + \alpha_{y_\lambda 3}k_3 - u_2^0 = 0 \\ \alpha_{\varphi 1}\delta\varphi + \alpha_{\sigma 1}\delta\sigma_\varphi + w_1 = 0 \\ \alpha_{\varphi 2}\delta\varphi + \alpha_{\sigma 2}\delta\sigma_\varphi + \alpha_{x_\lambda 2}\delta x_\lambda + \alpha_{y_\lambda 2}\delta y_\lambda + w_2 = 0 \\ \alpha_{x_\lambda 3}\delta x_\lambda + \alpha_{y_\lambda 3}\delta y_\lambda + w_3 = 0 \end{cases} \tag{2.29}$$

取初值$\sigma_\varphi^0 = \sigma_\varepsilon, \varphi^0 = \theta, x_\lambda^0 = x_i, y_\lambda^0 = y_i$，对方程组(2.29)迭代求解可得到$\delta\varphi$、$\delta\sigma_\varphi$和$\delta x_\lambda$、$\delta y_\lambda$。进而可求得曲线最大误差的方位角$\varphi_m$和$S_{\varphi\lambda}$，并以$S_{\varphi\lambda}$的最大值表示曲线不确定性的带宽，称之为$\varepsilon_m$模型，$\varepsilon_m$的轨迹所构成的曲线精度范围称为"$\varepsilon_m$-带"。

2.2.2　不确定性地理目标模型

基于模糊集和粗集理论的不确定性地理目标表达模型是现有不确定性地理目标模型中最为常用的方法。

1. 基于模糊度的不确定地理目标的形式化描述方法(何建华,2005)

基于模糊贴近度构建的不确定地理目标的形式化描述方法，对于不确定地理目标间的空间关系不确定性的进一步分析中有着非常重要的意义。

(1) 模糊点的形式化描述

点集拓扑中，点被认为是空间的基元，并以此基元分别定义了线和面对象。基于模糊贴近度的模糊点的形式化描述为：

$$\widetilde{P} = (x, y, A, D)$$

其中，x, y为点目标P的二维平面坐标，A为点目标P对应的属性集，

$A=\{a_1,a_2,\cdots,a_n\}$;D 为点 P 的模糊度。根据地理目标不确定性的不同来源,这里的 D 具有以下不同的物理意义。

对由于位置随机性引起地理目标的模糊性情形,D 理解为属性 A 位于点 $P(x,y)$ 的可能性程度。如由于量测误差带来的不确定性,假定其服从正态分布,则 D 可定义为:

$$D_1(x,y)=1-\frac{1}{2\pi\sigma_1\sigma_2}\exp\cdot\left\{-\frac{1}{2}\left[\frac{(x-x_0)^2}{\sigma_1^2}-\frac{2(x-x_0)(y-y_0)}{\sigma_1\sigma_2}+\frac{(y-y_0)^2}{\sigma_2^2}\right]\right\}$$

式中,x_0、y_0 分别为连续观测样本 x、y 的期望;σ_1、σ_2 分别为 x、y 的方差。

对由于主题属性的模糊性引起地理目标的模糊性情形,主题属性的模糊性主要分为分类的模糊性和概念的模糊性两类进行讨论,D 表达为点 $P(x,y)$ 隶属于属性 A 的模糊程度。

分类的模糊性,如遥感影像中的混合像元即是具有分类模糊性的像元,对它们的分类应根据专门领域的知识进行相关隶属函数的确定。例如,对林地的分类,遥感中经常使用的叶面积指数(LAI)则可作为这里的 D,即

$$D_2(x,y)=\sum S/A$$

其中,S 为 $P(x,y)$ 点所在分类单元单株树木叶面面积,A 为分类单元土地总面积。

概念的模糊性是模糊表达、分析和推理必须处理的另一大类问题,如距离的远近、质量的好坏等,对这类概念的模糊性不能给出一个准确的数值,而是一个范围。最常用的是三角模糊数的表达方式 (l,m,r),其中 l 为模糊集的上确界,r 为模糊集的下确界,m 为模糊集的核,相应的 p 点对应的 D 可定义为:

$$D_3(x,y)=\begin{cases}\dfrac{\|p-m\|}{\|m-l\|}, & p\in m-l;\\ 0, & p\in m;\\ \dfrac{\|p-m\|}{\|r-m\|}, & p\in r-m,\end{cases}$$

式中，$\|p - m\|$为点p与核m的某种范数，实际应用中通常取点p到点m的距离。

当地理目标同时带有位置的随机性和属性的模糊性甚至全部不确定性来源时，通过综合不确定性度量策略来确定地理目标的模糊度，从而集成表达不同来源的不确定性。利用合成运算方法（为取小、求差、求积、加权和等运算）作为合成函数f，最基本的方法是模糊集的交运算，即$D = D_1 \wedge D_2 \wedge D_3$。因此，模糊点目标的形式化描述为：

$$f_p(x_0, y_0, A_0, D) = \{\widetilde{p_1}, \widetilde{p_2}, \cdots, \widetilde{p_n}\}, 0 < D \leqslant 1$$

即模糊点是以确定点$P_0(x_0, y_0, A_0)$为核（$\mathrm{kern}(f_P) = P_0$），以$\widetilde{p_1}(x_i, y_i, A, D)(D > 0)$为支撑的模糊集合。

（2）模糊线的形式化描述

在确定的点集拓扑空间，线被定义为点的有序集。相应地，可定义模糊线为模糊点的有序集，即

$$f_L = \{f_{p_1}, f_{p_2}, \cdots, f_{p_n}\}$$

且满足：

$$\forall f_{p_i} \in f_L, i < n \rightarrow (f_{p_i}, f_{p_{i+1}}) \in f_L,$$

$$\forall i, j(i \neq j \wedge (i \neq 1 \vee j \neq n) \wedge (i \neq n \vee j \neq 1)) \rightarrow f_{p_i} \neq f_{p_j}$$

即相邻点构成线段，且除起点和终点外，没有两点相等。

相应地，模糊线即是以确定线$L = \{\mathrm{kern}(f_{p_1}), \mathrm{kern}(f_{p_2}), \cdots, \mathrm{kern}(f_{p_n})\}$为核，以$f_{p_1} \vee f_{p_2} \vee \cdots \vee f_{p_n}$为支撑的模糊集合。

（3）模糊面的形式化描述

在确定的拓扑空间，面被定义为由边界线所封闭的区域。相应地，在模糊拓扑空间，模糊面可定义为由模糊线封闭的区域，该区域是由模糊点组成的有限集f_R：

$$f_R = \{D(x, y, A)/(x, y, A)\}$$

其中，f_R为非空双连接闭集，f_R的内部为规则双连接开集，f_R的支撑等于其内部的闭包。

2. 不确定地理目标的粗集表达方法

粗集是用知识库中已知的知识近似表示未知的概念，主要通过

在未知概念与知识库中的基本知识之间建立一种关系函数(隶属函数)来实现。普通粗集的关系函数是两个集合之间的简单运算——交、并或包含,未知概念的下近似集合由所有包含在概念中的基本知识块的并集组成,未知概念的上近似集合由所有与概念相交的基本知识块的并集组成。将某个划分函数作用于论域集合并将其划分为人类所能识别的分类集,基本知识就是论域集合上的分类集组成的集合。图2.8为用粗集表达和近似概念的过程。

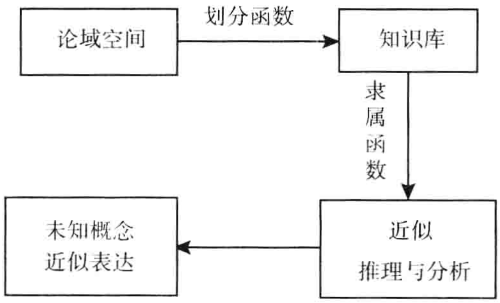


图2.8　用粗集表达和近似概念的过程

(1) 模糊目标的粗集表达方法

粗集是一种不确定性数据分析工具,能够对空间对象进行粒度分析。利用粗集对模糊目标进行处理不失为一个好方法。模糊目标的每个像元除带有灰度值外,还带有一个模糊隶属度值,表明它属于模糊目标的隶属程度。带有隶属度的像元不属于一个对象,它隶属于多个目标,从而完成对模糊目标模糊性的描述。模糊目标中像元的隶属度一般从遥感影像模糊分类结果中获取,也可以从这些不确定性或模糊分类中抽取模糊对象(Cheng Tao, 2001)。

设 U 为2维元素 (x,y) 的论域空间,A 为2维模糊对象,$u(x,y)$ 为它的隶属函数,对于任意的实数 $0<\beta\leqslant\alpha\leqslant1$,定义模糊目标 A 的下粗近似集 $\underline{A}_\alpha$ 和上粗近似集 $\overline{A}_\beta$ 分别为:

$$\underline{A}_\alpha = \{(x,y)\in U \mid u(x,y)\geqslant\alpha\}$$

$$\overline{A}_\beta = \{(x,y)\in U \mid u(x,y)\geqslant\beta\}$$

$\underline{A}_\alpha$ 为肯定属于 A 的隶属程度不小于 α 的元素的集合，$\overline{A}_\beta$ 为可能属于 A 的隶属程度不小于 β 元素组成的集合（张文修等，2001）。

模糊目标 A 的边界定义为：

$$\mathrm{bn}_R = \overline{A}_\beta - \underline{A}_\alpha$$

α，β 的取值直接影响着表达结果，α 取值越大，β 取值越小，对象的边界越大；反之，对象的边界越小。

当 A 为精确对象时，$u(x,y)$ 取值为 0 或 1，令 $\beta=0$，$\alpha=1$，则 2 维模糊集 A 的下粗、上粗近似集分别与 A 相等，即 $\underline{A}_\alpha=\overline{A}_\beta=A$，$\mathrm{bn}_R(A)=0$，这与 A 是精确对象的概念是一致的。因而，模糊目标和精确目标可以用粗集在同一个框架中表达，即任一空间对象 A 都可以用其下粗近似和上粗近似集来统一表达，这对研究模糊对象和粗糙对象的空间关系、处理和分析有很大的好处（杜世宏等，2003）。

（2）随机目标的粗集表达方法

传统的地图学中将现实世界视为一个离散空间，地图由形状各异的空间目标组成。而且，空间目标被认为是具有明确边界的点、线、面或其组合的表达（邓敏等，2006）。这些具有明确边界的空间目标在被研究其位置不确定性时，被称为随机目标，随机目标中点、线、面的位置不确定性分别用误差椭圆、误差带、误差环来表达。从误差椭圆、误差带、误差环的现实意义来看，随机目标的真实位置是不可知的，但是可以将目标界定在一个区域范围内。在这个区域内，随机目标的边界是不可分辨的。因此，用粗集方法来表达随机目标是可行的。首先将随机目标转换成栅格格式，然后用数学形态学的侵蚀和膨胀算子对其进行操作，得到基于误差椭圆、“ε-带（环）”，随机目标 X 的下近似、上近似和边界区域可分别表达为：

$$\begin{cases} \underline{X} = X \Theta \varepsilon \\ \overline{X} = X \cdot \varepsilon \\ \mathrm{bn}(X) = \overline{X} - \underline{X} \end{cases} \tag{2.30}$$

式（2.30）中，Θ 和 $\cdot$ 分别为目标形态侵蚀和膨胀算子。对于随机

点、线和面目标，它们的粗集表达如图 2.9 所示。

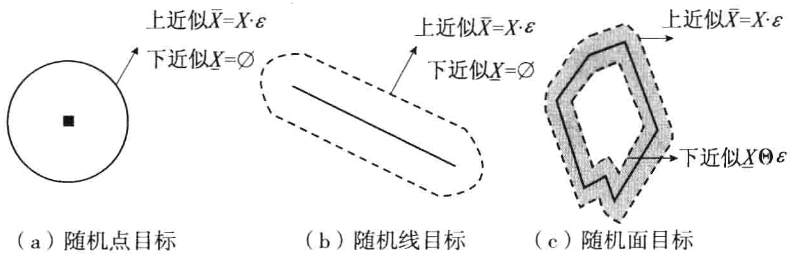


图 2.9　随机目标位置不确定性的粗集表达

分析可知，随机点和随机线目标的下近似为空集，即表明根据已有知识不能确定点、线的真实位置(在图2.9(a)和(b)中分别用灰色点和线表示)。而边界区的大小仅与 ε 的取值有关，也就是说，不可分辨的区域大小取决于获取的空间数据的精度。由此可知，对于大比例尺地图，空间数据精度高，不确定性带(即边界区)窄，数据的观测粒度小；而对于小比例尺地图，空间数据精度低，不确定性带宽，数据的观测粒度大。

2.3　本章小结

本章首先对本书处理现实世界随机不确定性和模糊不确定性所涉及的四种数学理论进行阐述，这四种数学理论分别是概率论与数理统计、信息论与熵、模糊理论、粗集理论；然后对已有的空间数据不确定性模型进行论述和分析，这些不确定性模型包括：点位中误差与误差椭圆、线元与曲线误差带(“ε_σ-带”模型、最大方向的误差带“ε_m-带”模型)、面元误差环等、基于模糊理论的不确定地理目标的形式化描述方法、随机目标的粗集表达方法等。这些数学理论与不确定性模型是构成本书研究空间数据多尺度表达不确定性的方法和理论基础。

第3章　空间数据多尺度表达不确定性及其形式化描述

研究空间数据多尺度表达不确定性的首要任务应该是在其不确定性来源分析和分类研究的基础上，建立不同尺度空间数据不确定性的统一的形式化描述方法，作为开展相关研究的基础。

本章3.1节对空间数据多尺度表达的实现方案进行分析对比，3.2节对空间数据多尺度表达不确定性的来源进行论述和分析，3.3节分析和确定空间数据多尺度表达不确定性的类型，3.4节建立空间数据多尺度表达不确定性的概念模型，3.5节构建空间数据多尺度表达不确定性的形式化描述模型。

3.1　空间数据多尺度表达的实现方案

多尺度空间数据库可认为是数据表达在空间域上多层次版本的集成。一方面要获得连续的尺度表达效果，另一方面又不能让数据量无限增大。在尺度变化空间，与尺度 s 对应的表达为 $f(s)$，由尺度 s_i 到 s_{i+1} 的表达变化为 $\Delta f_i = f(s_{i+1}) - f(s_i)$。从数据组织角度来看，针对数据量压缩和尺度变化粒度精细(在几何细节层次上变化)的特点，以数据库的存储内容为考察对象，空间数据多尺度表达的实现技术有四种策略(艾廷华等，2005)，第一种方法是多级尺度显示存储型(目前在实际应用中普遍采用的方法)，第二种方法是初级尺度变化累积型(顾及尺度特征的空间表达的新型数据模型)，第三种方法是关键尺度函数演变型，第四种方法是初级尺度自动综合型(比较理想的然而也是最难以实现的方法)。

1. 多级尺度显示存储型

多级尺度显示存储型是在数据库中储存来自同一个现实世界实体的多种表示，即建立对应于多种比例尺的多个数据库，如图3.1所示，数据库存储的数据集为$\{f(s_0), f(s_1), f(s_2), \cdots, f(s_i), \cdots\}$。其核心是根据GIS中对地理信息的不同需求层次，划分出一系列比例尺空间数据，来模拟空间现象变化的连续性。空间数据库是通过采集不同比例尺的空间数据来建立的，图形显示通过控制当时屏幕比例尺的变化，轮换调入和释放相应比例尺的数据来实现，应用时则提取某比例尺对应的空间数据集即可。这是目前国内外在实际应用中普遍采用的方法，如我国建立的1∶5万、1∶25万、1∶50万和1∶100万国家基础地理信息系统。法国国家地理院(IGN)制图署也采取这种方法开发了三个主要数据库：1∶2.5万的BD Topo、1∶5万的BD Carto和1∶100万的BD。在一些商用GIS软件和计算机图形学中也常采取这一方案，如ArcGIS、MapInfo等GIS软件。

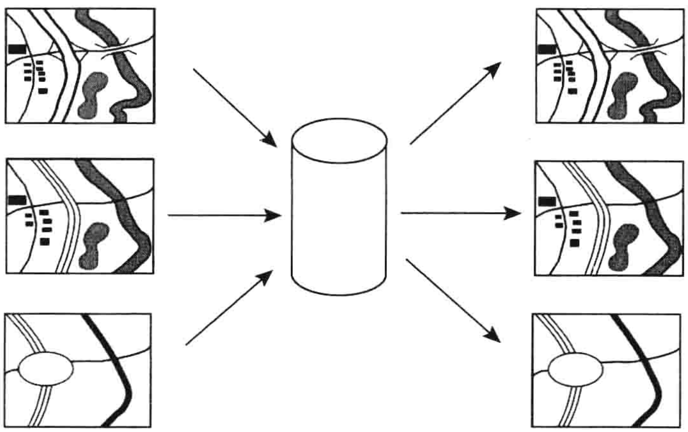


图3.1 多尺度多版本建立多重表示模型

该方法在一定程度上缓解了当前系统对不同详细程度数据的迫切需求和自动综合相对落后之间的矛盾，对早期GIS的发展和应用

起到了很大的推动作用。这种方法在不同比例尺上表示的同一地理实体的空间对象之间没有任何联系，各尺度之间没有平稳的过渡关系，一般不考虑多比例尺空间对象间的一致性和可维护性，技术完备一点的在不同比例尺下目标表达间通过目标匹配建立对应关系。因此，该方法数据冗余最大，尺度变化离散(不是真正的连续性尺度变化)，多比例尺上图形层次明显不连续和不协调，数据更新困难，耗费巨大。

2. *初级尺度变化累积型*

初级尺度变化累积型是顾及相邻尺度间的数据表达存在大量的相同的事实，提取存储相邻尺度间的变化部分，数据库存储的数据集为 $\{f(s_0), \Delta f_1, \Delta f_2, \Delta f_3, \Delta f_4, \cdots\}$ 。在对国家序列比例尺地形图 1∶1 万、1∶5 万、1∶10 万、1∶25 万、1∶100 万的同一区域的表达作比较时会发现，数据表达的变化是缓慢的，“表达生命周期”长的目标在多个比例尺段的图上均出现，而且没有表达形式上的明显差异。设在尺度空间 $\{s_0, s_1, s_2, \cdots, s_i, s_{i+1}, \cdots\}$ ，相邻尺度间的表达变化为 $\Delta f_i = f(s_{i+1}) - f(s_i)$， 则通过递推关系，可计算出任意尺度 s_i 下的表达为：$f(s_i) = f(s_0) + \Delta f_1 + \Delta f_2 + \cdots + \Delta f_i$，该公式表明，任意尺度下的表达可通过变化的累积叠加来实现，因此把它称作“初级尺度变化累积模型”，这是顾及尺度特征的空间表达的新型数据模型。这里的初级尺度对应小比例尺低分辨率下的粗糙表达状态。比例尺越大，累积的变化部分越多(上述表达式包含的项数越多)，对目标的表达越完备、真实。根据尺度变化粒度划分的层次，变化的内容可以是目标实体，也可以是几何细节单元。

该策略实现的关键技术包括：①相邻尺度表达间“变化”的提取；②“变化”组合的线性结构的建立；③“变化”绑定实现一定“粒度”的划分(该绑定过程相当于在上述表达式中合适位置加括号)；④“变化”累积后真实表达的恢复。根据尺度变化粒度划分的层次，变化的内容可以是目标实体，也可以是几何细节单元。基于目标实体的变化累积，实际例子中有基于 Horton 编码对水系流域的各河流目标分级，按比例尺建立线性索引，实现流域网络结构的多尺度

表达。在地图综合算法现有的成果中，基于层次结构的目标表达如GAP-tree，可以直接支持变化累积模型的建立。实现基于目标实体的变化累积模型相对要简单一些，但变化的粒度太粗糙。基于几何细节的变化累积要"细腻"，关键是要寻找有效的在尺度特征影响下目标与几何细节间的层次分解与组合，分解的基本单元可看作是前一个尺度与后一个尺度间的变化部分。

3. 关键尺度函数演变型

这是在第二种方法的基础上提出的，基于地物目标在比例尺表达空间上存在值的突变和量的缓变，数据组织时只存储记录发生质变的数据状态，中间的量变状态通过函数内插导出，从而大大降低了数据量。

该方法的空间数据库存储的数据集为 $\{f(s_{k0}), f(s_{k1}), f(s_{k2}), \cdots\}$，其中 s_{ki} 为关键尺度。空间数据库包括多个层次、详细程度逐渐增加的不同尺度(比例尺)的空间数据，高层的某一目标可能由多个低层目标来表达，从而构成层次数据结构，如图3.2所示。这种方法依赖层次数据结构的实现，即所有空间目标按其尺度来排序。通过这种多分辨率的层次数据结构表达空间数据，可以存储和检索不同详细程度的内容。空间数据集可以随比例尺的变化

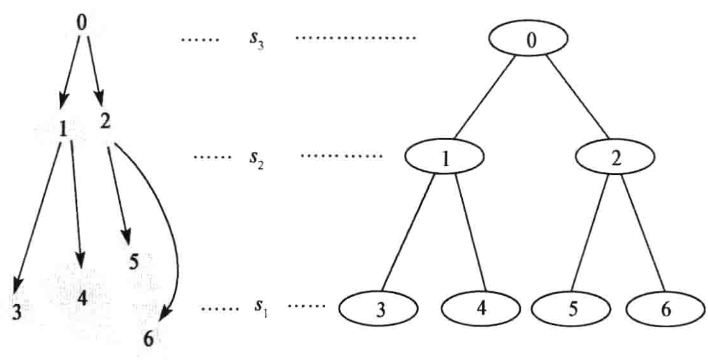


图3.2 多尺度对象的识别和联接及其层次树结构(有向非循环图)(李霖等，2005)

而增加或减少，但只能在该数据集允许的尺度范围内变化。关键尺度的确定取决于利用该数据集完成什么样的任务。Van Oosterom (1995)利用 Reactive-Tree(反作用树)技术，在开放式的面向对象数据库管理系统 Postgres 中建立了一个交互式的多尺度 GIS—CEO++。Timpf(1998)在其博士论文中基于图论也建立了这样的实验系统 DAG。这是目前认为比较可行的方法，但其数据库中各比例尺(层次)的空间目标不能自动地彼此相联，而这种联接对于识别在各层次上表达同一地理实体的空间对象是十分关键的。目前，基于人工预置的识别方法缺乏动态可维护性，因而层次树难以自动建立和维护。而且随着版本数量的增加，其生成层次并建立联接的过程会越来越复杂，效率也会越来越低。

4. *初级尺度自动综合型*

初级尺度自动综合型只存储初级尺度下的数据(单一的，有较高详细程度的主导数据库)，即数据库存储的数据集为 $\{f(s_0)\}$，任何其他较概略的表示由全自动化综合技术在初级尺度数据基础上按需要实时综合输出，如图 3.3 所示。这种方法是应用综合运算来产生数据库的多重版本，因而其关键之处是需要自动综合的能力，常规的人工制图综合使这一方法的模拟得以实现。这种 GIS 数据库能在不同等级上表示实体，并能确定实体间的联接关系，支持对数据的不同详细级别的修改。当一个实体被修改时，其变化能通过联接关系传播到其他实体，从而自动演绎出其他分辨率级别的数据，实现多重表示。这种综合方式就是所谓的 on-the-fly 综合。

但是，这一方案对数据库技术要求很高，数据库操作的工作量很大，尤其是比例尺跨度比较大时，其工作效率并不一定比重新产生相应比例尺数据的效率高，而且也可能造成很多错误。就目前的数据库技术和 GIS 技术还不足以实现该方案。这是一种比较理想的然而也是最难以实现的方法，是地图学家努力的方向(Timpf, 1998；王家耀，2001)。目前，数据综合只有部分工作易于实现自动化操作，如地物选取、线地物简化和光滑、面状对象合并等(Kilpelainen，2000)，其他操作则依靠人机交互综合方式来完成。

目前，比较公认的数据自动综合的观点是将综合过程分为两个

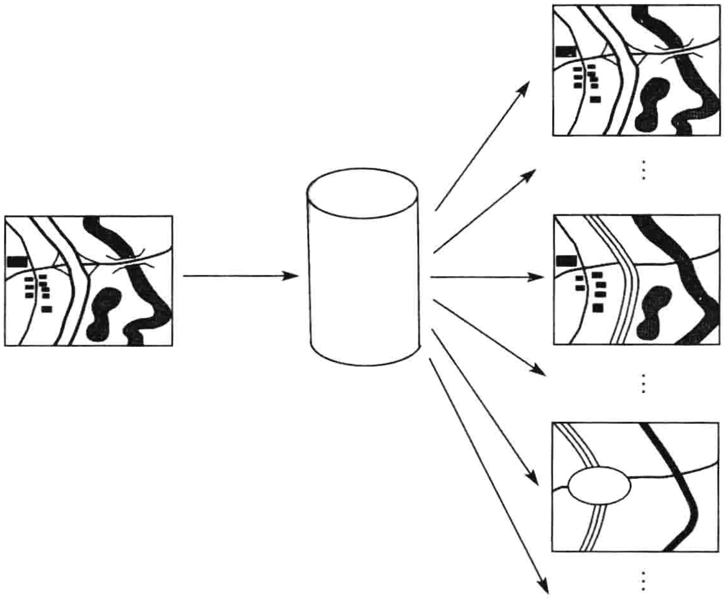


图 3.3　基于主导数据库建立多重表示模型

独立的阶段(Li Zhilin 等，1995；高文秀，2002)：

第一阶段是数据库的综合(database generalization)，也称为概念综合阶段(conceptual generalization)或语义综合(semantic generalization)或模型综合(model generalization)，从高分辨率或大比例尺数据库派生出低分辨率或小比例尺数据库，该阶段并不考虑图形显示方面的问题。

第二阶段是视图综合阶段(view generalization)，也称为地图综合阶段(cartographic generalization)或图形综合阶段(graphic generalization)，是数据的重新配置、重新调整，使数据能够在缩小的空间内合理地摆放或排列，解决数据可视化表达中空间对象之间矛盾冲突，具体讲就是利用几何图形简化、光滑算法，连同数据符号化手段使数据表达符合地图规范。图 3.4 是 Li Zhilin 等人(1995)提出的综合方案。

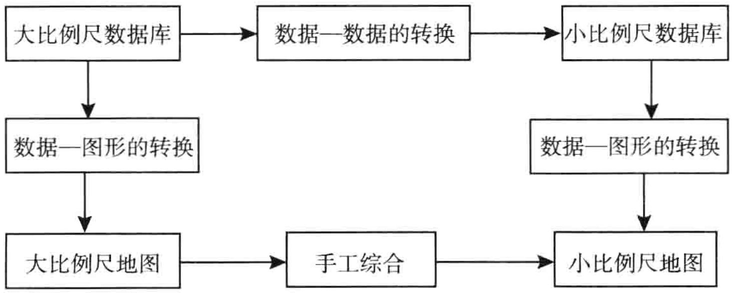


图 3.4　GIS 数据综合方案(Li Zhilin 等，1995)

3.2　空间数据多尺度表达的不确定性来源

空间数据的多尺度表达是在不同尺度(比例尺)下以相适应的不同详细程度表达的空间数据。在 GIS 实际应用中，有效实现空间数据的多尺度表达及多尺度表示数据的动态连接的主要途径是数据的自动综合，即初级尺度空间数据经过多次处理、转换后生成的产品。因此，初级尺度空间数据的不确定性以及各种处理转换过程中的不确定性必然传播给了多尺度表达的数据。空间数据多尺度表达的不确定性一方面来源于初级尺度空间数据的不确定性，另一方面来源于自动综合(各种处理转化过程)引入的不确定性，这两方面不确定性的综合影响作用即为空间数据多尺度表达的不确定性。

3.2.1　初级尺度空间数据的不确定性

初级尺度空间数据通过直接对地理空间进行抽象或空间采样来获得。现有研究中的空间数据就是这里的初级尺度的空间数据。

空间数据的不确定性包括可度量和不可度量的误差，也包括数据值和概念上的误差，其来源是非常复杂的，不同的学者从不同的角度进行了归类和总结，Bur rough(1986)将空间数据误差来源归为三类：明显的误差来源，由自然变化或原始测量引起的误差，GIS 处理过程中引入的误差。Goodchild(1988)将误差源分为：源误

差和处理误差。Maffini 等人(1989)将 GIS 数据误差产生的原因概括为三种：自然界的固有属性(变化性和模糊性)，测量的固有属性(观测误差的不可避免性)，表达测量结果的数据模型(地理模型结构的复杂性和近似性)。Herzog(1989)将误差源分为四类：数据采集误差、数据记录误差、数据处理和分析误差、数据图形化和可视化误差。Shi(1994)将遥感空间数据中的误差源划分为数据获取误差、数据预处理误差、数据转换误差和人工解译误差。刘文宝(1995)将 GIS 空间数据不确定性的来源分为五类：数据固有的不确定性、数据输入的不确定性、数据存储的不确定性、数据处理的不确定性以及数据输出的不确定性。承继成等人(2003)将空间数据不确定性的来源分为三大类，即客观世界固有的不确定性、人为引起的不确定性、地理空间数据的其他不确定性问题。史文中(2005)将空间数据不确定性的来源分为四个方面：客观世界本身的不确定性；人类认知过程的局限性及不确定性；空间数据量测误差；空间数据分析处理所引起的不确定性，并用图 3.5 直观地表示这四个方面的不确定性的来源及其影响过程。

下面分别对客观世界本身的不确定性、人类认知的不确定性、空间数据获取误差、空间数据分析与处理的不确定性进行详细阐述。

1. 客观世界的不确定性

客观世界是个非线性以及多参数的复杂巨系统。客观世界的复杂性是造成客观世界不确定性的主要原因之一。在客观世界中，相对于确定的实体而言，具有不确定或不均衡性的实体所占比例更大。某些空间实体没有明显的边界，或无法确定其边界，而许多实体目标的边界表现出渐变的特征，例如，土壤边界、森林或草地边界。非但空间实体边界的划分存在模糊性，即使被认为同一类的空间实体，其边界内部的物质结构也存在不确定性，不可能处处均值，而通常是异质的。客观世界的众多领域，如天文、地理、环境、地域划分、生物、气候、气象等领域都充满了不确定性。

在天文研究领域，我们现在所能“看到”的宇宙星体，其中的部分星体可能已经不复存在，而是已经消失了数千年甚至数万年。

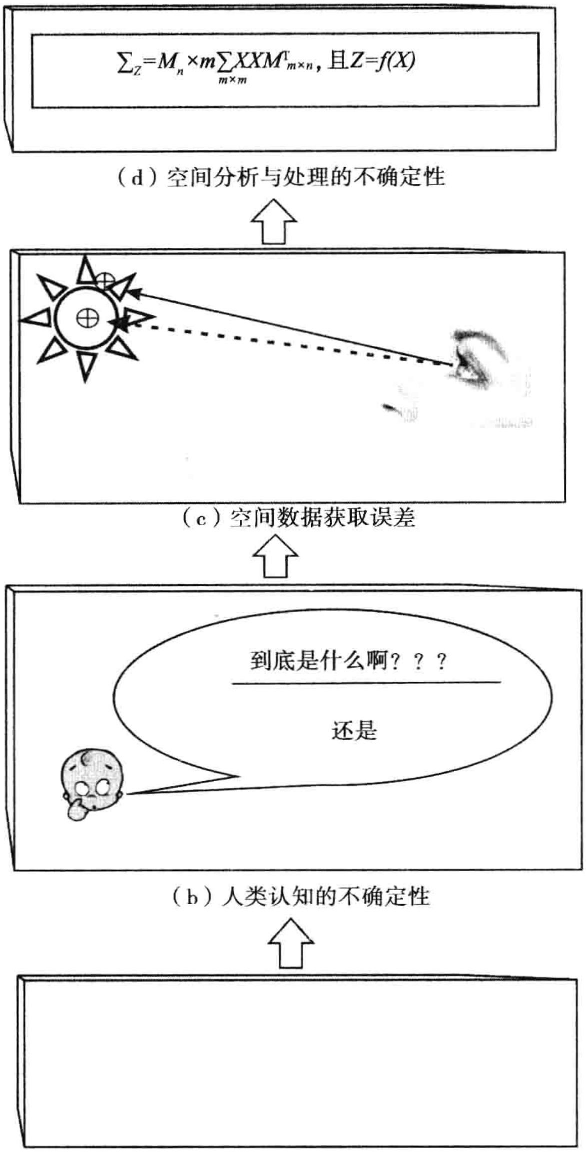


图3.5 空间数据不确定性的主要来源(史文中，2005)

而部分已经存在数千年、数万年的星体，我们却无法“看到”，可能要在数千年、数万年后才能“看到”它们。由于宇宙星际物质距离地球非常遥远，它们与我们的距离是用光年来计算的，因此当我们能够“看到”它们的时候，这些星际物质呈现出的状态可能是在若干年前的状态，现在其位置已经发生了变化。因此，我们从天文摄影相片所看到的宇宙是不确定的。此外，具有“波粒二象性”的光也并不总是像人们想象的那样按照直线运行，在“黑洞”附近，光可能发生弯曲，改变其原来的运行方向。因此，天文领域以“光年”量度宇宙星际物质间距离也具有明显的不确定性特征。综上可知，天文领域充满了不确定性的特征。

在地理研究领域，人们总是假设 GIS(在传统的 GIS 设计中)所处理的对象在自然界有清晰的界线。目标模型总是假设空间实体是由具有精确已知坐标的点，连接着系列精确已知点的线，以及由确定的线所严格界定的面域来描述的。然而，大多情况下，客观世界中具有几何定义的精确的点、线、多边形并不存在(Wang S. L. 等，2001)。例如，自然界不同属性的土壤和植被之间，其物质成分和结构物质是连续的、逐渐变化的，不一定存在明显的分界线。各个自然地理带之间，如不同的气候带、不同的植物带之间都具有渐变的、逐步过渡的特征，大多数情况下不存在明显的分界线。又如水域岸线，一般认为地理空间上的水域边界线是三维的。然而，按 Mandelbrot 的观点：从分形来看，像香港的海岸线这样的曲线，具有分数维数，且不为唯一，这与传统的精确几何学的概念是相冲突的。可见，海岸线具有不确定性的特征。实际上，所谓的地理界线大多是人为划定的，而它所对应的自然界地理状况则是不确定的。

在环境领域，不同性质的气体、水体之间相互渗透，其界面是动态的、千变万化的，不存在明确的界线。土壤风蚀、雨水冲刷区的界线也是不明确的、逐渐过渡的。不同气候带之间不同属性的植被也是渐变的，没有明确的界线。事实上，客观世界中，与环境相关的各类空间实体间均不存在明显的界线。所谓的界线都是人为界定的，环境中具有明显不确定性特征。

在地域划分中，城市与乡村之间的划分具有明显的不确定性。城市的中心区、边缘区，城市的边界，区域的边界(如珠江三角区、泛珠江三角区等)也具有明显地域或经济地域的不确定性。各地域之间都是渐变的，逐步过渡的，大多不存在明确划分的几何意义上的界线。

在生物领域，新陈代谢是生命的固有特征，物种的进化与退化，遗传基因的变异以及种群数量的增减都具有明显的不确定性特征，不确定性是生物领域的普遍现象。从达尔文的进化论角度看，生物的进化动力主要是由"遗传"、"变异"和"选择"构成。"遗传"使得生物种群得以保持，它既是生物进化论的基础，又是生物种群确定性特征的保证；在一个物种的成员间必须存在"变异"。"变异"是个体之间区别的基础，它为进化创造了条件，没有"变异"，便没有进化；"选择"决定了生物的进化，即"优胜劣汰，适者生存"，自然"选择"对"遗传"和"变异"发生作用，导致新物种的形成或旧物种的灭绝。经过长时间的"遗传"、"变异"和"选择"，生物逐渐从简单到复杂、从低级到高级不断进化。生物进化不是完全按照事先确定的模式进展，而是充满了不确定性，受到许多随机因素的影响。生物的进化是熵减，是不确定性原理一般性的客观体现。

在气象、气候上，气象要素变化无常。虽然气象预报的研究历史久远、投入较多、设备极为先进，但预报结果越来越充满了不确定性。以美国为例，100个气象站的预报结果为：5天的温度预报准确率为75%，一个月的温度预报准确率为61%，一个季度的温度预报准确率为58%。而降雨的预报准确率更低，5天的为59%，1个月的为52%，一个季度的为51%。我国运用每秒可计算数千亿次的计算机，根据大气状况，通过求解天气演变的方程组预报未来的天气，其准确率仅为60%～70%。这些实例足以说明气象问题本身存在很大的不确定性——不论我们怎么去努力，100%的预报准确率是难以达到的。气象预报是一个十分不确定的问题。

综上所述，客观世界的各个领域都充满了不确定性。

2. 人类认知的不确定性

认知是人脑反映客观世界及其物质的特征和联系，揭露客观世界及其物质对人的意义和作用的心理活动，即个体理解和获得知识的过程。认知能力和认知功能的形成及其方式随个体年龄和经验的不同而发生变化，并且每个个体都具有自己的认知风格。因而个体在知觉、记忆和思维模式上都存在着个别差异。总体来说，认知能力取决于社会生产力发展水平、个体受教育程度、个体经验的丰富度和掌握信息量的多少。与客观世界的复杂性相比，人类的认知能力非常有限。

人类社会是在不断发展变化的，人类对客观世界的认知也是逐渐深化的。人类对宇宙和地球的认识经历了漫长的过程。从最初具有神话色彩的“天宫”到各类行星、恒星的陆续发现，从对地球和太阳系的认知到浩瀚无际的河外星系的认知，每一次认知无不打上了人类科技进步的烙印。对宇宙认知的每一次进步，揭示了前一次对宇宙认知的不确定性和局限性。不幸的是，人类对于宇宙的认知和探索往往受到反科学势力的影响。历史上，在人类社会中占统治地位的神权意志曾经压制了科学的发展。人类对宇宙认知变得逡巡不前。众所周知，从基于神权统治的“地心说”的崩解到基于科学的“日心说”的普遍认可，其间是付出了惨重的代价的，这种人为的影响无疑加深了人类对宇宙认知的不确定性，也延缓了人类认知宇宙的进程(承继成等，2003)。

人类对地球的认知也是一个漫长的过程，每一个认知阶段都充满了人类对地球认知上的不确定性——从“天圆地方”发展到现在的认知水平大约耗去了1000年的时间。从《山海经》到今天的中国地图付出了许多代中国人的艰辛。在气象预报方面，从“天有不测风云”到天气预报准确率达70%用了数百年的时间。地球的地壳结构、地质构造等认识也经历了逐步深化的过程——“地台地槽学说”、“地洼学说”、“大陆漂移学说”、“海底扩张学说”和“板块构造学说”——每一种学说的出现都代表了人们对地球的又一种认知，也说明了地壳的复杂性和认知的艰难性。现在“板块构造学说”又遭到了人们的质疑，人类对地壳的认知仍然具有相当的不确定

定性。

人类自诞生以来，已经与周围的生存环境相处了100多万年，但至今人类对环境的认知仍然处于初级阶段。“环境决定论”、“人定胜天论”、“人与环境合一论”等仍然作为“人与环境关系”的争论热点，至今没能完全统一。可见，人们对环境的认知有了一定的进步，然而每一步的发展都是艰难的。现在，又有一种新的理论——“生态理论”出现了，这足以看出人类对环境乃至客观世界的认知过程总是在不断地发展。

关于地理认知的研究始于20世纪80年代。它以地理学、心理学和认知心理学为基础，并广泛涉及制图学、城市规划、环境心理学、语言学、哲学以及人工智能等其他相关领域的学科。其研究内容主要包括人类有关地理时空的领悟力、记忆力、推理能力、目标属性组合及其表达、各种事件之间的相互通信能力等。一般地，地理认知研究主要包括两类：地理空间认知研究，地理信息认知研究（龚建华等，2002）。有关地理空间认知的研究，已经被美国地理科学研究协会于1996年列为“地理信息科学在未来十年内的十大优先研究领域”之一。从国外研究来看，目前有关地理空间认知的研究主要包括：地理空间作为一个有关人的“心理空间”或“经验空间”是怎样变化的；人们是怎样获取地理信息的；个体的年龄、文化、性别或特殊的背景等因素是如何影响人们对于地理信息的认知的（鲁学军等，1999）。

与客观世界的复杂性相比，人类的认知能力是非常有限的。为了满足生活的需求和欲望，人们试图尽可能地利用其能够理解并熟练操作的技术来感知其身外的客观世界。首先，人们感知客观世界中的空间实体，观测此空间实体，了解、定义并归纳此空间实体，将反映空间实体属性的观测资料表达成为一种确定的概念数据模型，以正规术语来描述其属性数据，并将其结果存储于计算设备中。其次，基于GIS数据库中的存储数据，人们得以编辑、重复获取和分析其对应的空间实体。当GIS用户实施操作时，如属性特征的合并，其他类别的属性特征就有可能聚合成某种特定的特征。

由于客观世界中空间实体具有复杂性和可变性，人们必须选择

最为重要的空间特征来近似逼近真实实体。所有的属性数据均是借助于某些理论、技术和方法来获取的。这些理论、技术和方法隐含地或明确地指出了属性数据所需的抽象和概括的必要水平。因此，对于空间实体的属性而言，我们所描述的数据远少于客观总体数据。属性数据的期望值是由感知的目标近似逼近的。实际上，期望值也受制于当今人们认知的局限性，且空间数据库也仅仅是对自然界的一种抽象的表达方式。因此，经计算机处理而表达过的空间实体可能会丢失相应空间实体某些方面的特征，这就使得空间数据库中存在空间实体的不确定性。

而且，人们对客观世界的认知是在一定的尺度下进行的。认知的深度随着尺度的变化而变化。因此，在获取空间实体信息时，人们就要慎重选择适当的尺度，因为这直接决定了所获取的空间实体真实信息的不确定度。例如，运用不同分辨率的观测技术对同一物体进行观测所得到的观测结果可能是完全不同的，用微米、纳米观测的世界和用肉眼看到的世界是完全不同的，此所谓一种尺度，一个世界。同样的，在地形测量中，对于同一测区，运用大比例尺测绘的地图比采用小比例尺测绘的地图内容丰富、负载信息多，其不确定度相对较小，完整性较高。

综上可知，客观世界的复杂性和人类认知的局限性是造成人类认知不确定性的两个主要原因。

3. 空间数据获取误差

空间数据是描述地球的基础信息，是 GIS 的“血液”。基于空间数据的 GIS 已广泛应用于社会各行各业，如防汛减灾、城市规划、银行选址、气象预报等。空间数据与社会经济发展直接相关。因此，空间数据的获取显得十分重要。

一般来讲，广义的空间数据获取技术包含空间数据获取方法的选择、仪器与人员的优化配置以及技术方案的最优化设计等方面。空间数据获取的主要方法有数据转换、内插、地图数字化、摄影测量直接获取数据、常规测绘技术和 GPS 测量技术、遥感技术等。然而，由于社会生产力发展水平和人类认知能力的限制，空间数据获取技术存在其局限性。空间数据获取技术的局限性必然导致所获

取的空间数据存在不确定性。空间数据获取技术的局限性主要体现在量测设备固有的精度范围、量测技术及其实施方案、空间数据获取技术的现势性等方面。

即使具有先进的量测设备和科学合理的优化技术方案，在空间数据获取中，空间实体本身存在的时空上的不确定仍然会导致空间数据获取结果具有不确定性。

空间数据获取误差是空间数据不确定性的重要来源之一。这里，误差的概念指的是测量学领域经典的数据处理的误差概念，包含数据获取过程中引入的系统误差、随机误差和粗差。

对客观世界中空间实体的测量而言，“测得准”与“测不准”并存是空间实体的又一特征。理论上，空间实体都有属于某属性的真值。但是，由于测量技术的限制，仅能获得表达真值的量测近似值，而不可能获得真值。工程建设中，桥梁中轴线的长度及其定位精度、隧道的长度及其贯通精度、GPS 静态控制测量精度等一般是较高的。但是，这些观测值也并非真值，同样存在误差，只不过这些误差处于规范容许的范围之内。无论什么时候，利用何种设备和技术，量测值没有误差的情况是不存在的。因此，量测设备与技术的不确定性加大了人们对客观世界认知的不确定性。

4. 空间数据分析与处理的不确定性

由空间数据获取技术所获取的数据为原始数据，必须经过“加工”——空间数据分析与处理，之后才能被用户接受和使用。广义上的空间数据分析与处理，指的是对所获取的空间及非空间数据的分析与处理，包括诸如数据的存储、更新、转换、查询、统计处理以及输入、输出等。对每一种空间数据处理而言，或多或少地会引入不确定性，并且会继承数据本身的不确定性或误差。

(1)空间分析的不确定性

空间分析是分析、模拟、预测和调控地理事件及空间过程的一系列理论和技术，其分析结果依赖于事件的空间分布，并面向最终用户。利用 GIS 空间分析功能，通过对原始数据模型的观察和实验，用户可以获得新的知识和发现，并以此作为空间行为的决策依据。然而，由于空间数据总是受到不同类型不确定性的影响，而这

些不确定性又通过空间分析而传播，其结果势必导致空间分析的结论不精确。

(2)计算设备引入的不确定性

根据人们的需求，当计算机设备中存储的GIS数据用于描述、获取、存储、操作和分析空间实体时，这些数据都是通过二进制形式来表示客观世界的(Ganter，1988)，这是对连续的自然界的一种近似表达和处理。某些不确定性可能来自计算设备本身，如物理建模、逻辑建模、数据编码、数据操作、数据处理、算法优化、计算设备的精度以及输出精度等。另外，个别属性的实际值与编码之间总是存在差异。

即使设计和操作是完美的，任何测量设备也只能以一定的精度记录其测量成果。即使利用具有足够高精度的测量设备获取数据，如果使用低精度的输出设备(如显示器、打印机等)，则其输出成果可能依然缺乏足够高的精度表达。对于某些应用，通常通过选用更为精确的测量设备，以降低成果中的不确定性。然而，设备的优选也仅能近似逼近客观实体，而无法达到绝对准确。具有无限高精度的真实测量成果是不存在的，所有测量成果都具有不同程度的不确定性。

(3)运算过程带来的不确定性

利用计算机实施空间数据分析处理时，我们往往要通过建立模型和优化算法来进行。其间，模型的选择、算法的优化、数据的质量及其数量等都对分析结果带来不同程度的不确定性影响，这些不确定性主要有如下体现：

①在建立各类模型或算法过程中由于认知的不完整性、模型的不确定性、算法的不可靠性而导致整个运算过程及结果的不确定性特征。

②量测原始数据、公式参数难以获得真值而引起的不确定性。

③模型固有的不确定性。

④数据质量欠佳、数据量不完整、数据密度不够大而导致的运算不确定性。

⑤凡是通过模型计算获得而又未经验证或无法验证的数据，都

有可能具有不确定性特征。

(4)异源数据融合带来的不确定性

将来源不同、可靠性各异的异源属性数据实施融合时，其结果的不确定性变得更为复杂。为了创建一种最为理想的数据库，GIS 数据用户可能会需要诸如某类均值或多源要素组合等异源数据的匹配和融合。但是，一般的空间数据库多支持精确的局部的分析而对全局性的分析应用显得不足。假如按照全局性分析应用的要求将不同的局部数据库进行融合，那么不同空间数据库之间的矛盾就会凸显出来。因此，除了上述源自客观世界、人类认知、数据获取、计算设备或处理技术等不确定性外，异源数据的融合是空间数据的另一不确定性来源。

总之，GIS 数据的不确定性是不可避免的。在 GIS 分析过程中，当若干不确定性得到累积时，GIS 数据不确定性可能传播甚至通过累积而变得更大。同时，人们的认知水平、数学模型和处理技术的局限性有可能进一步扩大空间数据的不确定性。

3. 2. 2　自动综合引入的不确定性

实现数据多尺度管理和表达的主要途径是进行 GIS 数据自动综合，地理信息的自动综合是空间数据多尺度表达的核心内容(王家耀，2001)。

多尺度的空间数据与初级尺度的空间数据相比较存在的差异，就是由自动综合引入的不确定性，因为多尺度的空间数据是初级尺度的空间数据自动综合的结果。自动综合分为模型综合和地图综合(图形综合)两个阶段，因而自动综合的不确定性分为模型综合的不确定性和地图综合的不确定性两大类。

模型综合是从初级尺度(大比例尺)的空间数据(库)经过信息抽象派生出多级尺度(小比例尺)的空间数据(库)，并不考虑图形显示方面的问题。一方面，模型综合是空间数据库中数据的选取，选取保留下来的数据没有发生任何变化，其空间位置精度得以保持；另一方面，模型综合的结果是对应于同一地理景观或地理目标的数据减少了，也就是地理内容的复杂程度降低了，意味着空间信息的减少，也就是说，模型综合的目的就是为了减少空间信息。信

息论认为信息是不确定性的减少量，空间信息减少，不确定性必然增多。

地图综合(图形综合)是在模型综合结果的基础上，进行数据的重新配置、重新调整，使数据能够在缩小的空间内合理地摆放或排列，解决数据可视化表达中空间对象之间的矛盾冲突，具体地讲就是利用几何图形简化、光滑算法，连同数据符号化手段使数据表达符合地图规范。从综合过程和综合结果来看，地图综合过程在一定程度上损失了地理实体的位置精度，但是可以较好地描述地理实体的空间形态(Muller，1996)。

以上是从理想的自动综合的目的和过程的分析来看的，自动综合必然导致空间数据多尺度表达相对于初级尺度空间数据的不确定性。

但是，从现实的自动综合算法或系统来看，自动综合引起的空间数据多尺度表达的不确定性的原因远比理想的自动综合复杂。一是因为自动综合是常规的人工地图制图综合方法的模拟实现，但它无法再现人脑创造性的混沌的思维过程，不能根据地理目标或地理景观的特点自动调整参数或算法程序，使得具体的自动综合算法实用对象受到限制，对其不适应的地理目标或地理景观甚至会得到错误的结果。二是因为自动综合算法的设计实现不是针对具体地理目标或地理景观的，同一地理目标或地理景观可用不同的自动综合算法来综合，算法不一样，结果也就不一样，引入的不确定性也有差别。三是因为对于不同的比例尺，应该综合到什么程度，没有统一的标准，自动综合算法也不能进行不同比例尺间的综合变换，综合程度如何至今仍未找到合理的、数量化的评估方法和标准(Joao，1998；郭庆胜，1998)。四是因为自动综合是在一定的计算和显示设备上进行的，设备精度不一样，结果也会有所差异。

王家耀(2004)将自动综合划分为一系列有序算子的集合：数据变换、分类分级、选取、化简、合并、图形等级变换、典型化、夸大、注记选取配置、关系处理、拼接、符号化，自动综合是这一系列有序算子的联合，各个算子之间应既有联系，又有相对的独立性，既要考虑算子的需求性，还要考虑实现的可能性，完备的自动综合算子集合要涵盖整个自动综合过程。这 12 个自动综合算子的

综合应用是一个非常复杂的过程，每一个算子的实现都会引入不确定性，这些不确定性一方面受人的认知水平的影响，另一方面受软硬件设备的影响。

因此，无论是从理想的自动综合机制还是现实的自动综合模型、自动综合算子来分析，自动综合都为多尺度表示的空间数据引入了不确定性。就是常规的人工地图制图综合方法也不可避免地为多比例尺的地图引入不确定性。因为制图综合本身就是对初级尺度地图的抽象，抽象必然导致地图信息的减少，而且这种抽象还受地图用途、比例尺、景观条件、图解限制和地图数据质量的影响，使得综合结果存在不确定性。另外，由于受制图者的认知和技术熟练程度的影响，对于同样的地图、同样的要求，不同的制图者综合的结果不一样，即使是同一制图者，多次综合的结果也有所差异。

由于自动综合会引入空间数据的不确定性，而且不同的算法引入的不确定性也有所差异，因此，已有若干对自动综合算法的质量评价，郭庆胜(2000)研究了地图自动综合的评价标准，见表3.1。

表3.1　**地图自动综合的评价标准**

观点	以地理分析为目的的综合	以地图生产(或可视化)为目的的综合
几何	精度范围内误差最小，空间关系不变	误差在给定条件下最小，空间关系不变
艺术	用户易操作，视觉效果好	符合地图生产的整饰规范
传输	有利于提高地理分析的效率，保留地理分析所需的地理特征	能准确反映出地理特征，易读
技术	计算效率高(能满足用户要求)	动态可视化时响应时间很短，生产地图时经济效益最优
表示法	同地理分析的抽象程度相适应(在空间目标维数、属性精度等方面)	符合地图表示法的原则

他将地图自动综合分为以地理分析为目的的综合和以地图生产(或可视化)为目的的综合，综合结果的评价标准涉及几何、艺术、传输、技术、表示法五个方面。比如对于线地物的自动化简算法评价中，Douglas-peuker方法具有严格的保凸凹性，另外保留弯曲的深度指标，使它有严格的弯曲的质量控制，但在某些不适当的深度指标下，它具有自相交的情况，计算工作量大；步行法综合曲线不能保持弯曲的特征；光栏法算法严格，能按给定阈值保留曲线的特征点，而且能进行实时处理，运算量小，占用内存少；小波方法可以获得自适应的综合表示和空间特征的无级表达等。由此可见，对自动综合算法的质量评价，其实就是对自动综合引入多尺度空间数据不确定性的分析和评价。

3.3 空间数据多尺度表达的不确定性类型

3.3.1 概述

不确定性是数据的主要特征之一，因此，空间数据都存在不确定性，根据第1章图1.1空间数据的分类，空间数据不确定性的分类如图3.6所示。

现有研究中，属性数据、空间数据、时域数据的不确定性都有了大量研究成果。由于本书所研究的空间数据是狭义的空间数据，也就是地理空间数据中的空间数据，或者称为图形数据，不包含属性数据和时域数据。因而，空间数据的多尺度表达，实际上是图形数据的多尺度表达。多尺度表达的空间数据实际上是空间数据的派生数据，因此，空间数据多尺度表达的不确定性，其实就是不同尺度上空间数据的不确定性，也包括几何数据的不确定性和关系数据的不确定性。

在点、线、面数据不确定性的研究中，点数据主要讨论其位置不确定性，线数据主要讨论其位置不确定性和形状不确定性，面数据主要讨论位置不确定性和面积不确定性。

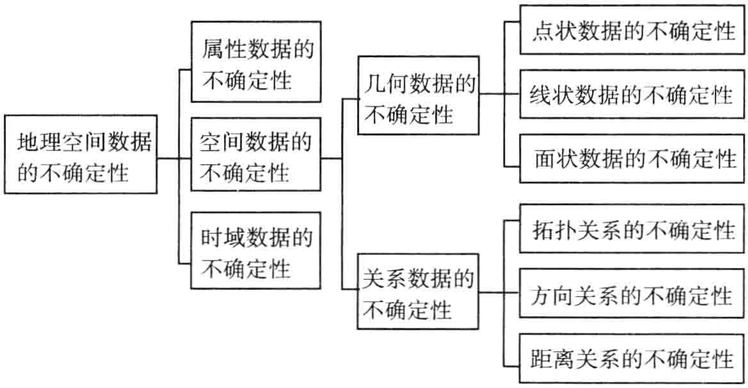


图3.6 空间数据不确定性的分类

在空间数据库中，不直接存储关系数据，而是通过几何数据直接计算出来的，在现有关于空间关系的研究中，主要讨论拓扑关系和方向关系的不确定性问题。

因此，在本书的空间数据多尺度表达的不确定性研究中，主要研究空间数据多尺度表达的位置不确定性和空间关系(拓扑关系、方向关系)的不确定性。

3.3.2 几何不确定性

GIS中的地物目标(object)大体上可分为天然目标和人工目标。人工目标，如一个体育场，一栋房子、一条铁路，它们的位置、长度或面积是确定性的，不同的人用不同的仪器去量测(统计取样)会得到不同的位置、长度或面积量测值。从理论上讲无穷次量测将能得到其真值，有限次量测是对真值的估值，因而是确定性目标。自然目标，如海岸线、河流岸线、火山口的边界、天然林与草地的边界等，它们有分形特征，表现为不满足处处连续光滑和各向同性，而呈现出处处不可微分的特征，因而是不确定性目标。

对于确定性目标，其位置、长度或面积的真值是客观存在的，从理论上讲是可以测量出来的，因而，确定性目标的几何不确定性

是其量测值与真值的差值，叫做真误差或误差。现有研究中将真误差分为系统误差、偶然误差和粗差。由于系统误差可以减弱或消除，粗差可以剔除或避免，而偶然误差是不可避免的，因而对于确定性目标几何不确定性的研究主要是研究其偶然误差，现有研究中称为随机误差，主要用概率论与数理统计方法进行研究，常用中误差、方差等指标来表示。在测绘科学中，针对确定性地理目标的几何不确定性，尤其是位置不确定性已有大量的研究成果，如点位中误差、误差椭圆等，线元误差带，面位误差环等。

对于不确定性目标，在地理空间中的几何位置只有一个模糊的范围，现有研究中称其为模糊不确定性。现有研究中除了近似地用统计数学方法处理外，还用分形数学、模糊数学、粗集理论和集统计与模糊数学为一体的云模型理论等数学方法来处理。

无论是确定性地理目标，还是不确定性地理目标，其对应的几何数据在尺度转换中都会发生变化，因此，多尺度空间数据的几何数据的不确定性与初级尺度的不确定性是不相等的。到目前为止，关于多尺度空间数据的几何不确定性的研究成果非常少，本书第4章将针对几种典型的空间数据多尺度表达的几何不确定性进行分析和建模。

3.3.3 空间关系不确定性

GIS空间关系主要分为顺序关系、度量关系、拓扑关系三大类。由于地理目标自身的不确定性，空间数据的获取、处理的不确定性以及人类对空间关系的认知的不确定性，必然导致地理空间目标之间空间关系的不确定性。在空间数据库中，空间关系是利用某空间关系表达模型通过空间数据直接计算出来的，空间数据存在不确定性，空间关系也就必然存在不确定性。空间关系的不确定性可以从现实世界、认知世界、GIS(数字)世界3个方面来认识，如图3.7所示(杜世宏、王桥，2004)。在现实世界中，空间关系的变化是连续的、渐变的，而且既有量的变化，又有质的变化，其空间关系的种类繁多，约束复杂。现实世界中的空间关系给GIS提供了研究空间关系的原型。

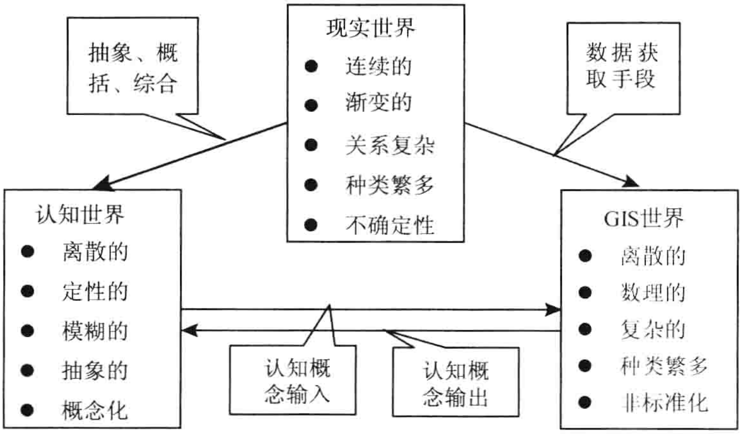


图 3.7　空间关系在 3 个世界观中的特点(杜世宏、王桥，2004)

在认知世界中，空间关系是人们通过对客观现实世界中空间关系的抽象、概括、综合以及总结归纳后产生的结果。由于人们受认知能力的限制，无法对现实世界中的所有空间关系的各个方面进行建模和表达，因此，在认知过程中，人们需要利用概括、综合和抽象等手段来简化空间关系。但是，由于抽象、概括等，认知世界中一个空间关系概念对应着客观现实世界中多个实例，客观现实世界中一个实例对应着认知世界的多个概念，也就是认知世界中的空间关系概念集合与现实世界中的实例集合之间是多对多的映射关系，而这种映射关系则具有某种不确定性。

综上所述，由于空间关系在客观现实世界、人类认知世界和 GIS 世界中的特点和表现方式不同，而且总是存在差异，因此产生了不确定性。由于 GIS 世界中空间关系是通过空间数据计算出来的，因此，空间关系的不确定性主要是由认知的不确定性、数据的不确定性以及空间关系分析处理的不确定性等引起的。

由人类认知引起的空间关系的不确定性主要表现为空间关系概念的模糊性。由于 GIS 不但是对客观现实世界的建模，同时又为人类认识和改造现实世界服务，因此 GIS 必须能够以人们所希望的方式来处理所管理的数据。假设给每个空间关系概念定义一个函数

$\phi(O_1, O_2)$，即可建立概念与空间对象 O_1 和 O_2 之间的映射关系。在已有的空间定性关系表达中，$\phi(O_1, O_2)$ 的值域为0(不能用该术语表述)或1(能用该术语表述)，因此一对对象间空间关系只能用一个术语表述。因此，定性表达是一种等价划分，且各个等价类概念之间没有重叠，其函数值域是{0，1}，但是，人们的认知具有模糊性，这种模糊性是在认知客观世界的思维中形成的。由于同一组对象间的空间关系可以用多个概念来表示，且有不同的隶属度，因此其函数 $\phi(O_1, O_2)$ 的值是连续的，可在区间[0，1]内取值。这是一种模糊划分，且模糊等价类别(术语)之间有一定的重叠度，如图3.8中的3个对象，虽都可以用线在面内来描述，但是隶属度却是递减的。

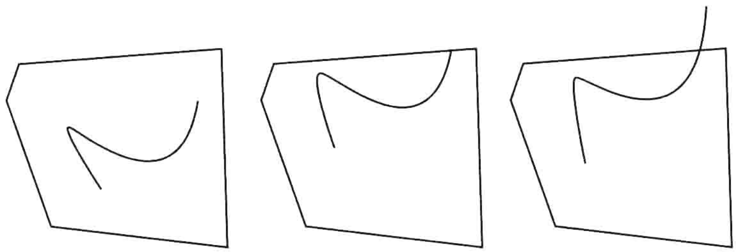


图3.8　拓扑关系的认知模糊性

人们认知的另一个显著特征就是在谓词前增加修饰语，如“差不多”、“大概”、“基本”、“完全”等，又如“对象A完全在对象B的内部”、“对象A与对象B基本相交”等语句。这些都可以归结为空间关系认知的不确定性，它们是通过概念的模糊性来表现的。

由数据不确定性引起的空间关系不确定性，主要是指由确定性地理目标边界位置误差和不确定性地理目标边界附近属性的模糊性导致的空间关系不确定性。通常，确定性点的不确定性用误差椭圆来描述，确定性线的不确定性用“ε-误差带”模型来描述，面对象的边界也不再是几何意义上的线，而是具有一定宽度的带状区域，从而使得空间关系在边界地区发生变化。不确定性目标其边界属性具有模糊性，如城乡结合部、土壤类型分布等。这些现象一般采用

场模型来表达，场中的每个像元均以不同的隶属度隶属于多个对象（分类），这便造成了模糊对象空间关系的不确定性问题。

由空间关系分析处理中的不确定性引起的空间关系的不确定性，主要是指空间关系分析处理的模糊性引起的空间关系的不确定性，通常空间关系在 GIS 中的应用包括数据检索、数据匹配、关系推理、相似性及完整性判断等，这些运算都可能涉及模糊性。基于空间关系的数据处理一般可以归结为空间关系相似性问题，并有精确相似与模糊相似之分，其中，空间关系精确相似只有相似与不相似两种状态，而没有相似的程度，这是不合逻辑的，而模糊相似性却可以给出两个空间关系的相似性程度。若用函数 $S(r_1, r_2)$ 表示空间关系 r_1 和 r_2 之间的一致性程度，则对于精确相似来讲，$S(r_1, r_2)$ 的值域是 $\{0, 1\}$，而对于模糊相似来讲，$S(r_1, r_2)$ 的值域是 $[0, 1]$。另外，还可以在模糊相似性定义的基础上，做出进一步的模糊定义，如“空间关系 r_1 与空间关系 r_2 基本相似”、“空间关系 r_1 与空间关系 r_2 完全相似”等。

由于空间关系包括拓扑关系、方向关系和距离关系，因此，空间关系的不确定性包括空间拓扑关系的不确定性、空间方向关系的不确定性和空间距离关系的不确定性。现有研究中针对空间拓扑关系和方向关系的不确定性较多，而很少有空间距离关系的研究。

通过以上的分析得出，由地理目标的几何数据计算出的空间关系，其不确定性一方面受几何数据不确定性的影响，另一方面受人们认知和空间关系分析处理方法不确定性的影响。相应地，对于多尺度空间数据，与之对应的空间关系因为受到不同于初级尺度空间数据的多尺度空间数据的几何不确定性的影响，因此，多尺度空间关系的不确定性也不同于初级尺度的空间关系。

3.4　空间数据多尺度表达的不确定性概念模型

3.4.1　概述

不确定性问题存在于空间数据的生命周期全过程，只要有空间

数据就可能有不确定性存在。空间数据的整个生命周期过程，也就是空间数据不确定性的生命周期过程。艾廷华(2010)根据空间数据在尺度上展示出的不同形式的演变，以尺度为主线，建立了尺度空间中GIS数据表达的生命期模型，该模型按照比例尺从大到小的演变顺序，构建尺度的线性空间，该空间中实体目标被表达的序列表征集 $\sum_{s_0}^{s_n} R_s$ 或连续的表达状态变化 $\int_{s_1}^{s_2} R_s$ 构成了GIS数据在其生命期的表达。其中，R_s 为尺度s处的表达，$[s_0, s_n]$ 为生命期的范围。该模型认为尺度空间构成了GIS数据表达的生命期，因此，在所有尺度上的空间数据都存在不确定性。

基于空间数据生命周期也是空间数据不确定性生命周期的思想，下面以尺度空间中空间数据不确定性的产生、建模、表达和控制给出空间数据多尺度表达不确定性研究的概念模型。

尺度空间中空间数据表达的生命期内的每一个环节都有不确定性产生，如何认识每个环节中不确定性产生的原因和机理，并用具体的模型表达且进行不确定性评价，是空间数据多尺度表达不确定性研究首要的也是核心的问题。空间数据多尺度表达不确定性研究的最终目的，是对其不确定性进行控制，所有这些都最终服务于空间数据多尺度表达的质量。空间数据多尺度表达不确定性概念框图如图3.9所示。

①空间数据多尺度表达不确定性来源：该单元给出了空间数据多尺度表达不确定性产生、来源，包括尺度转换过程中的不确定性，并给以简单的描述；

②空间数据多尺度表达不确定性建模：该单元主要包括尺度转换过程中不确定性模型建立的一般方法和步骤；

③空间数据多尺度表达不确定性表达：该单元包括空间数据多尺度表达不确定性的产生、传递，不确定性评价结果的图表、文字及可视化表达；

④空间数据多尺度表达不确定性控制：该单元描述了空间数据多尺度表达不确定性控制的流程；

⑤空间数据多尺度表达质量控制：该单元描述空间数据多尺度

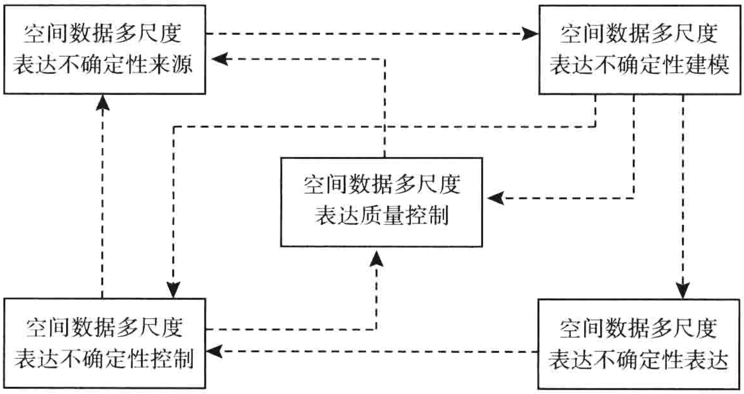


图3.9　空间数据多尺度表达不确定性的概念模型图

表达不确定性与空间数据多尺度表达质量间的关系以及提高空间数据多尺度表达质量的重要措施。

3.4.2　不确定性的传播

原始数据存在着不同程度的误差，经过各种处理、转换后生成的产品也保存着原有的误差，这就是不确定性的传播。空间数据的多尺度表达就是初级尺度的空间数据经过多次处理、转换后生产的产品，因此，初级尺度空间数据的误差以及各种处理转换过程中产生的误差必然传播给了多尺度表达的数据。当然，各种误差在传播过程中可能进行累积也可能发生相互抵消。因此，空间数据的多尺度表达的不确定性，是尺度空间中空间数据表达的生命期所有不确定性的综合结果。而且，用多尺度的空间数据进行空间关系分析处理时，多尺度空间数据的不确定性会传播该空间关系表达的结果。

为了有效地研究空间数据多尺度表达结果的几何(位置)不确定性，可以将空间数据的整个数据生命周期分为初级尺度的空间数据(数据采集、绘制地图、数字化、数据存储)、模型综合、图形综合、图形显示等阶段，分别对各阶段的不确定性加以研究评价，再综合考虑各阶段的不确定性来研究空间数据多尺度表达结果的不

确定性。由此可见，空间数据多尺度表达结果的几何不确定性是一个间接计算量，满足误差传播规律。

例如，通常情况下，空间数据库中的线状特征数据用一串有序的（x，y）坐标对表示，因此，线状特征数据多尺度表达的不确定性是由空间数据库中数据的误差与自动综合阶段产生的误差合成的，自动综合阶段的误差又是数据化简、曲线光滑、图形显示各误差合成的。假定各误差是不相关的，线状特征数据集多尺度表达的节点精度可用下面的公式表示：

$$\sigma^2 = \sigma_{ds}^2 + \sigma_s^2 + \sigma_{sim}^2 + d_{sim}^2 + \sigma_{sm}^2 + d_{sm}^2 + m^2\sigma_v^2 + m^2\sigma_{sb}^2 \tag{3.1}$$

式(3.1)中，σ_s 为数据的存储误差；σ_{ds} 为空间数据库中数据的误差；σ_{sim} 为数据化简时的计算误差；σ_{sm} 为曲线光滑时的计算误差；m 为多尺度表达的比例尺；σ_v 为多尺度表达自适应可视化的设备显示误差；σ_{sb} 为曲线符号表达误差；d_{sim} 表示数据化简时初始线上点到简化线上的最短距离（位移矢量）；d_{sm} 表示曲线光滑时弦弧间的偏差。以上是利用误差传播规律在假定各误差相互独立的条件下得到的线状特征数据多尺度表达的节点的不确定度计算公式。

从上面的分析中可以得出，多尺度的线状特征数据其不确定性大于初级尺度线状特征数据的不确定性。对于多尺度空间关系，在实际应用中由与之对应的多尺度空间数据派生计算出来，多尺度空间数据的不确定性必然会按照误差传播定律传播给对应的空间关系，因此，多尺度空间关系的不确定性可以按误差传播定律计算出来。当然，在现有的空间关系不确定性计算方法中，由于空间关系表达模型的复杂性，并不直接按照误差传播定律计算空间关系的不确定性，而是采用区间理论、模糊数学、粗糙集等数学方法来分析与表达空间关系的不确定性。

3.4.3　空间数据多尺度表达不确定性建模

在分析了空间数据多尺度表达不确定性的来源之后，就可以对尺度空间中空间数据表达的生命周期内各个环节内的不确定性进行分析并建立数学模型。建模时需要对不确定性的产生过程进行抽

象、简化，所以建模本身也有一定的不确定性。如何抽象、概化，以最大限度来模拟空间数据不确定性的产生过程，将直接影响空间数据多尺度表达不确定性的表达与控制。

空间数据多尺度表达不确定性建模问题主要有：空间数据多尺度表达位置不确定性建模，空间关系多尺度表达不确定性建模，空间数据多尺度表达属性不确定性建模等。各种模型的一般建立流程如图 3. 10 所示。

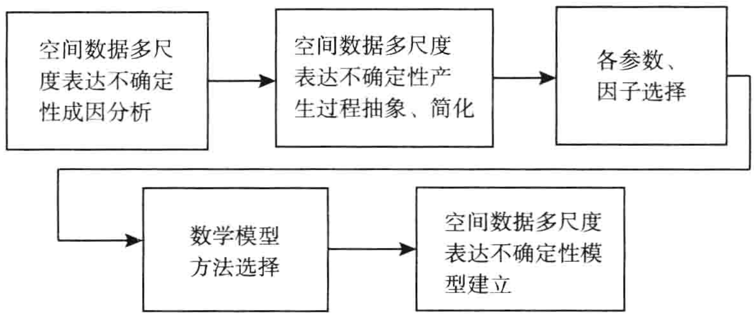


图 3. 10　空间数据多尺度表达不确定性模型的一般建立流程

3. 5　空间数据多尺度表达不确定性的形式化描述

本节探讨一种能集成表达空间数据、空间数据表达尺度及空间数据不确定性的形式化描述方法。

1. 空间数据多尺度表达的形式化描述

艾廷华，李精忠(2010)定义了尺度空间 GIS 数据表达的生命期模型，并定义了对象 O 在表达生命期内任意尺度下的表达模型，该模型作为空间数据多尺度表达的形式化描述是非常贴切的。

尺度空间 GIS 数据表达的生命期模型的构建思路为：

①借用生命期的概念，认为尺度空间构成了 GIS 数据表达的生命期。按照比例尺从大到小的演变顺序，构建尺度的线性空间，该

空间中实体目标被表达的序列表征集 $\sum_{s_0}^{s_n} R_s$ 或连续的表达状态变化 $\int_{s_1}^{s_2} R_s$ 构成了 GIS 数据在其生命期的表达。其中，R_s 为尺度 s 处的表达，$[s_0, s_n]$ 为生命期的范围。

②在生命期中，尺度 s_i 处的表达可以是预先存储好的数据版本，也可以通过尺度变换函数临时导出。在现有地图综合技术条件下，难以自动实现的或实现代价太高的表达，可通过基态显式存储，作为尺度空间中的关键性表达，在基态表达上通过尺度变换函数导出其他非关键尺度点上的表达。

③定义生命期内某目标表达的基态为 $\{g_0, g_1, g_2, g_3, \cdots, g_i, \cdots\}$，在其基础上定义三元组形式的尺度变换 T_i(i 表示该生命期上尺度变换的序列号，整个生命期内的尺度变换有多个)：

$$T_i: \langle f_i, [s_{i1}, s_{i2}], g_i \rangle$$

式中，$[s_{i1}, s_{i2}] \in [s_0, s_n]$ 为 T_i 尺度变换的作用域(s_{i1}、s_{i2} 为两个关键尺度点)，g_i 为该作用域内的表达基态，f_i 为作用于 $[s_{i1}, s_{i2}]$ 的尺度变换函数。

④对象 O 在表达生命期 $[s_0, s_n]$ 内，具有尺度变换 T_0，T_1，T_2，…，T_i，…，T_m，则该对象在任意尺度 x(x 位于且仅位于生命期的某子集尺度段内)下的表达为：

$$R_x = f_i(g_i, x), x \in [s_{i1}, s_{i2}] \tag{3.2}$$

式(3.2)即为空间数据多尺度表达的形式化描述。

2. 多尺度空间数据不确定性的形式化描述

空间数据整个生命周期过程，也就是空间数据不确定性的生命周期过程。按照尺度空间 GIS 数据表达的生命期模型的定义过程，同样可以定义尺度空间 GIS 数据不确定性表达的生命期模型。

空间目标在任意尺度状态下都存在不确定性，这里对于随机目标，其不确定性用“ε-带”模型描述，对于模糊目标，其不确定性用模糊度 D 描述。因此，对应于目标表达的基态 $\{g_0, g_1, g_2, g_3, \cdots, g_i, \cdots\}$，随机目标位置不确定性指标分别为 $\{\varepsilon_0, \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_i, \cdots\}$，模糊目标模糊度分别为 $\{D_0, D_1, D_2, \cdots,$

D_i，…}，在其基础上定义五元组形式的尺度变换下的不确定性变换 U_i（i 表示该生命期上尺度变换的序列号，整个生命期内的尺度变换有多个）：

$$U_i: \langle f_i, [s_{i1}, s_{i2}], g_i, F_i, \varepsilon_i / D_i \rangle \tag{3.3}$$

式(3.3)中，$[s_{i1}, s_{i2}] \in [s_0, s_n]$ 为 U_i 不确定变换的作用域（s_{i1}、s_{i2} 为两个关键尺度点），g_i 为该作用域内的表达基态，f_i 为作用于 $[s_{i1}, s_{i2}]$ 的尺度变换函数，F_i 为作用于 $[s_{i1}, s_{i2}]$ 的不确定性变换函数，ε_i、D_i 为该作用域内的表达基态 g_i 的不确定性指标。

对象 O 在表达生命期 $[s_0, s_n]$ 内，具有随尺度变换 T_0，T_1，T_2，…，T_i，…，T_m 的不确定性变换 U_0，U_1，U_2，…，U_i，…，U_m，则该对象在任意尺度 x（x 位于且仅位于生命期的某子集尺度段内）下的不确定性表达为：

$$\varepsilon_x / D_x = F_i(g_i, \varepsilon_i / D_i, x), \; x \in [s_{i1}, s_{i2}] \tag{3.4}$$

式(3.4)即为空间数据多尺度表达不确定性的形式化描述。

3. 空间数据多尺度表达及其不确定性的形式化描述

上面分别定义了空间数据多尺度表达和空间数据多尺度表达不确定性的形式化描述方法。将它们合并，得

$$(R_x, \varepsilon_x / D_x) = (f_i(g_i, x), F_i(g_i, \varepsilon_i / D_i, x)), \; x \in [s_{i1}, s_{i2}] \tag{3.5}$$

式(3.5)为生命期的某子集尺度段 $[s_{i1}, s_{i2}]$ 内任意尺度 x 下空间数据多尺度表达及其不确定性的形式化描述。该模型对于空间数据的基态只有初级尺度的情况下仍然适用，此时，尺度变换函数只有一个 f，不确定性变换函数只有一个 F，式(3.5)转化为：

$$(R_x, \varepsilon_x / D_x) = (f(g_0, x), F(g_0, \varepsilon_0 / D_0, x)), \; x \in [s_0, s_n] \tag{3.6}$$

3.6　本章小结

本章首先对空间数据多尺度表达的四种实现方案进行分析对比，这四种实现方案分别是：多级尺度显示存储型（目前在实际应用中普遍采用的方法），初级尺度变化累积型（顾及尺度特征的空

间表达的新型数据模型)，关键尺度函数演变型，初级尺度自动综合型(比较理想的然而也是最难以实现的方法)。其次，研究了空间数据多尺度表达不确定性的来源包括：初级尺度空间数据不确定性的传播、自动综合引入的不确定性。再次，对空间数据多尺度表达的不确定性进行了分类，并重点分析了空间数据多尺度表达的几何不确定性和空间关系不确定性。然后，建立了空间数据多尺度表达不确定性的概念模型，并重点研究了初级尺度空间数据不确定性在尺度转换过程中的传播规律和传播模式及空间数据多尺度表达不确定性的建模过程和方法。最后，构建了空间数据多尺度表达不确定性的形式化描述模型。空间数据多尺度表达不确定性的形式化描述方法是发展和建立空间数据多尺度表达各类不确定性评价模型的基础。

第4章　空间数据多尺度表达的不确定性建模

空间数据是地理信息的数字载体，是客观地理世界的抽象表达。空间数据的多尺度表达是对原始数据进行综合的结果。原始数据存在着不同程度的误差，经过综合后生成的多尺度产品也保存着原有的误差，而且综合过程中引入了新的误差并传播给了多尺度表达的数据。本章根据空间数据多尺度表达的生成机制，以及不确定性在这个过程中的传播机理，构建出空间数据多尺度表达不确定性的评价指标和量化模型。

4.1　点元多尺度表达的位置不确定性

点元 Q_0 在尺度 s_1，s_2，…，s_n 上表达的结果有三种可能，一是保留在原来的位置不动，二是移位，三是被删除。对于第一种情况，它在各尺度的不确定性与初始状态的不确定性是一致的。对于第三种情况，点被删除了，就不必要再考虑其不确定性了。对于第二种情况，点的位置不与初始尺度重合，这时点的位置不确定性一方面受其初始状态的位置不确定性的影响，另一方面要受位移矢量的影响。因此，在点位移的情况下，其不确定性为初始尺度不确定性与位移矢量的合成值。

若 $Q_i(x_i、y_i)$ 相对于 Q_0 点发生了位移，则 Q_i 相对于 Q_0 的位移矢量 d_i 在二维空间中为：

$$d_i = Q_i - Q_0 = (x_i - x,\ y_i - y)$$

$$|d_i| = \sqrt{(x_i - x)^2 + (y_i - y)^2}$$

$|d_i|$ 即为点移位引起的位置误差。由于点初始位置数据的获取与移位是相互独立的，因此，Q_i 的点位不确定性 σ_{Q_i} 可表达为一个基于误差传播定律的初始点位误差 σ_{Q_0} 和位移误差 d_i 的合成值，即

$$\sigma_{Q_i}^2 = \sigma_{Q_0}^2 + d_i^2 \tag{4.1}$$

由于点的初始位置以一定的概率落入误差椭圆内，因此，Q_i 的位置不确定性也可以表达为 Q_i 距离 Q_0 所在误差椭圆的最大值 $d_{i-\max}$，

$$d_{i-\max} = |d_i| + d_{\text{ellipse}}(x,\ y) \tag{4.2}$$

上式中，$d_{\text{ellipse}}(x,\ y)$ 为 Q_0 到 Q_iQ_0 延长线与误差椭圆交点的距离，如图 4.1 所示。根据 Q_0、Q_i 的位置可计算出 Q_iQ_0 直线的坐标方位角 θ，于是

$$d_{\text{ellipse}}(x,\ y) = \cos^2(\theta)\cdot\sigma_x^2 + \sin^2(\theta)\cdot\sigma_y^2 + \sin(2\theta)\cdot\sigma_{xy}$$

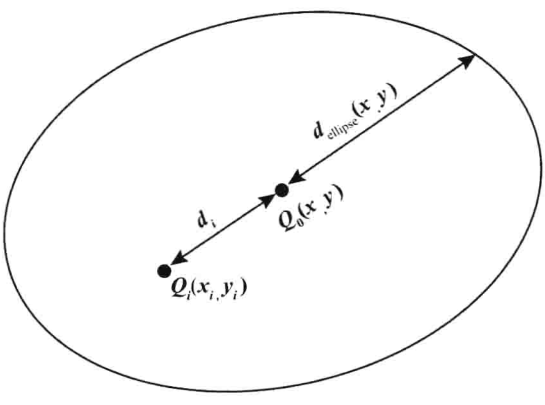


图 4.1 点位移与初始点的对比

σ_{Q_i} 和 $d_{i-\max}$ 均可作为点发生位移的不确定性度量指标，σ_{Q_i} 可作为 Q_i 的点位中误差，$d_{i-\max}$ 表示 Q_i 与 Q_0 的最大可能距离。若图 4.1 中点 Q_0 的点位中误差为 $\sigma_{Q_0} = \pm 35.2\text{m}$，利用式(4.1)和式(4.2)进行计算，得到 $\sigma_{Q_i} = \pm 37.9\text{m}$，$d_{i-\max} = 39.0\text{m}$。

4.2　线状要素多尺度表达的不确定性

在地图数据库中，线是较大的数据集。许多地理现象的线状面貌都能用抽象的几何线来表示，如铁路和公路的中心线、行政(州、市)区域和经济区域(属性、性能)边界、公共事业管线(电力线、电话线、自来水和污水管道)、自然现象的边界线(如植被、土壤)、海岸线等。Cuenin(1972)估计在地图数据库中大约有 80% 为线。而且通常线的空间分布比较均匀(如道路网络)。因为线是用节点及连接节点的线元定义的，所以线包含着大量的几何信息。

4.2.1　线状要素多尺度表达的方法

本书中线状要素的多尺度表达主要是指线状要素几何细节层次上的多尺度表达，即线地物的简化与光滑，这些都属于能够实现自动化操作的范围，且可以建立无比例尺的 GIS 数据库，存储内容可以尽量详细，即线状要素多尺度表达的实现策略为初级尺度自动综合型。

对于空间数据库中线状要素几何细节层次上的自动综合，对应于目前比较公认的数据自动综合的观点，可将其分为两个阶段完成，第一阶段是对无比例尺线状数据进行模型综合，采用各种线状数据的简化算法(如 Douglas-Peuker 算法等)，得到不同比例尺所需要的简化的线状数据；第二阶段是对简化的线状数据进行图形综合，采用曲线光滑(插值或拟合)函数对简化的线状数据进行光滑，得到满足需求的可视化曲线。线状要素自动综合的简单模式可用图 4.2 表示。

4.2.2　线状要素多尺度表达的不确定性

线状要素原始数据的获取以及为了实现多尺度表达所进行的自动综合，都在不同程度上引起了线几何特征的改变，包括线位置的改变和形状特征的改变，降低了线数据的可靠程度。如图 4.3 是对以相对弦长为参数的正轴抛物线加权平均法拟合表达的线状实验数

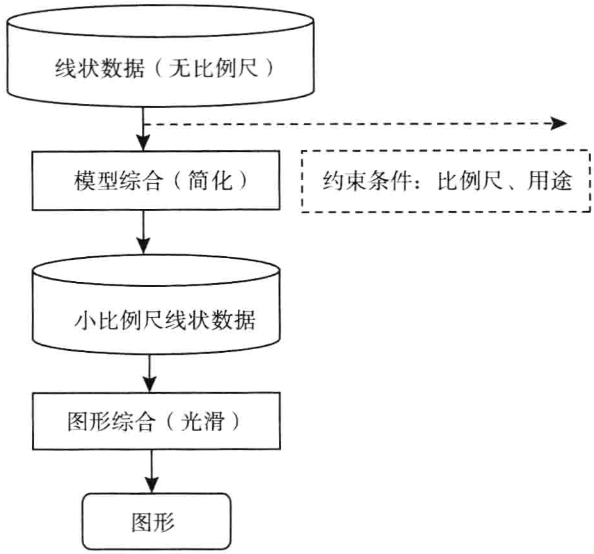


图 4.2 线状要素自动综合模式

据用小波变换进行多尺度表达的实例，它是对原始数据用 Daubechies 三阶小波进行 5 级尺度的分解，保留全部的近似系数，而舍去全部细节系数后，分别重建其结果，从而得到相应整数尺度上的多尺度近似表达，图 4.3 中 5 个近似数据与原始数据存在着差别，曲线的形状即曲线的复杂程度变得越来越概略。图 4.4 是原始数据与 5 个尺度上近似表达的重叠，随着曲线形状的化简，曲线的位置也发生了改变。

1. 位置不确定性

线状要素多尺度表达的实现流程是一般空间数据的流程的特例，可以用框图加以表达，如图 4.5 所示。

图 4.5 真实地展现了线状要素多尺度表达的实现流程，并且确立了各环节中的引入不确定性的因素。

由此可见，引起线状要素多尺度表达不确定性的原因除客观世界固有的不确定性外，还有量测误差（偏差、随机误差）、存储管

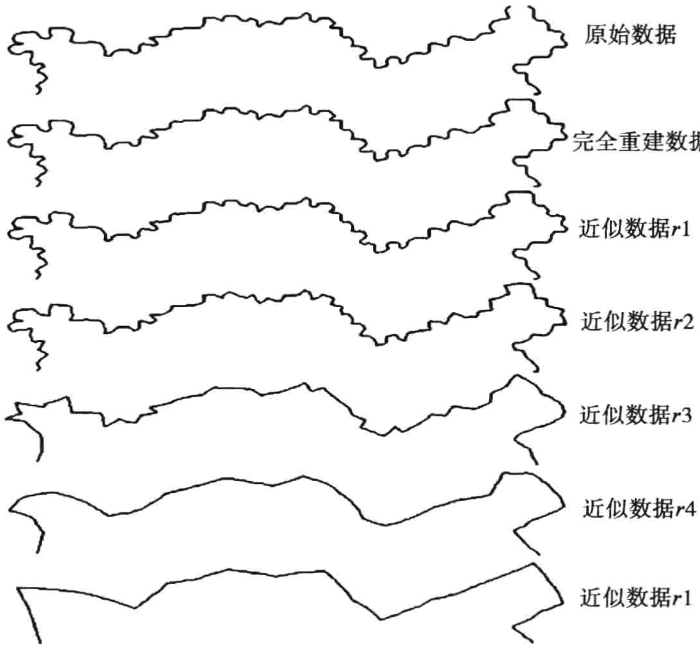


图 4.3　线状特征数据在 5 个尺度上的近似表达

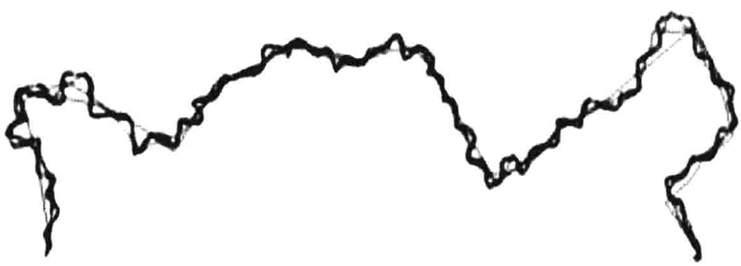


图 4.4　原始数据与其在 5 个尺度上近似表达的重叠

理中的误差、比例尺引起的不确定性、自动化简算法的不确定性、曲线光滑(拟合、插值)算法的不确定性、曲线表达符号的不确定性、可视化设备的不确定性等。

尽管线状要素多尺度表达的过程比较复杂，引发其不确定性的

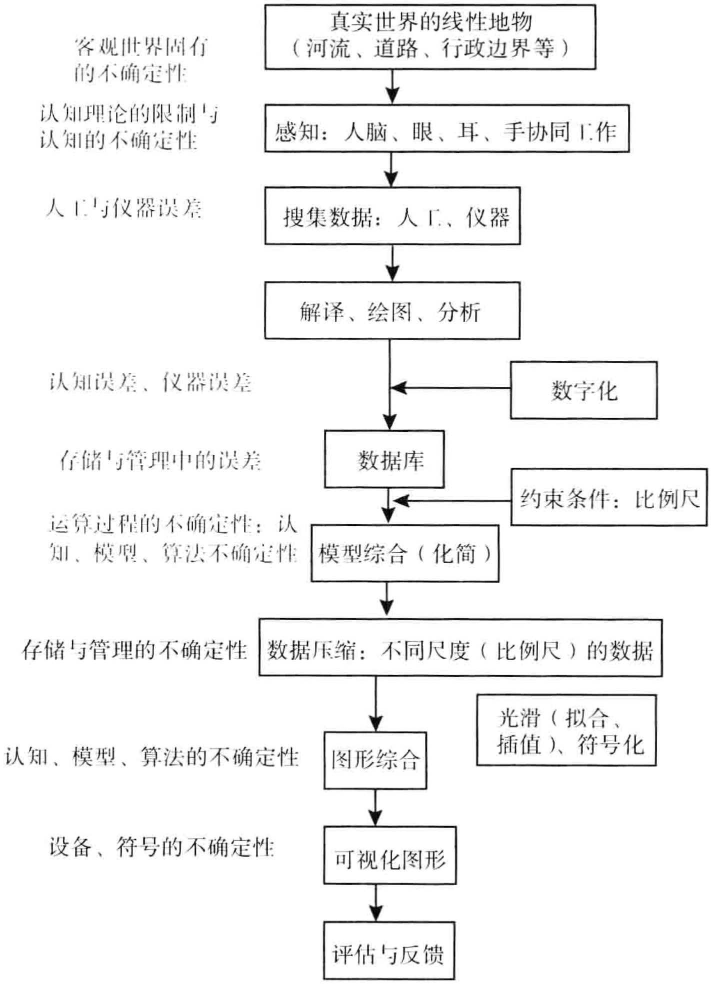


图 4.5　线状要素多尺度表达的实现流程

因素也很多。但在这个过程中，线基本上被定义为坐标串和线元。坐标串代表线上有明确意义的点。线元代表线，并能确保真实线和图形线的差异从表达比例尺度上考虑不至于过大。也就是说，节点是定义线结构的元素并且与真实线相关，线元是某一尺度最好的

表达。

因此，线状要素多尺度表达的位置不确定性表现为由节点位置不确定性及连接节点的线元的位置不确定性共同构成。

因为节点位置与线状要素真实位置对应，是真实位置对应的空间数据，所以，节点位置不确定性被定义为数据不确定性。连接节点的线元，其位置受曲线光滑模型的影响，光滑曲线方程不一样，曲线的位置也就有差别，所以，连接节点的线元的不确定性被定义为曲线光滑模型不确定性。也就是说，线状要素多尺度表达的不确定性由数据不确定性和模型不确定性共同构成。

2. 形状不确定性

线状要素多尺度表达通过自动综合技术而实现，实际上是对线状要素进行不同程度的化简而得到线状要素多比例尺的表达。进行化简的主要目的就是使相应比例尺对应的存储量最小，并保持线的弯曲特征。由此可见，在线状要素多尺度表达中，它的基本形状结构特征不会变化，只是几何细节有所变化。几何细节的不同程度的变化会导致线的几何特征发生不同程度的变化。

从图4.3可以看出，随着表达尺度的变化，曲线变得越来越概略，细节层次逐渐减少，长度、方向、轮廓都发生了不同程度的变化，但是它们存在一定的相似性。

在进行线的自动综合时，线的几何特征的描述是非常重要的(Buttenfield，1991)。Buttenfield(1991)把描述线的几何特征的规则分成三条：①随着比例尺的减小，地图上的线在符号化时的改变、删除、保留等方面会因要素不同，变化规律也会不同，有时是一种“突变”；②同一条线不同部分其分维数不同，应当建立步长与长度的关系；③建立基于尺度的参数规则，例如，结构特征(structure signature)——线段首末点连线的长度及其内部点偏离“连线”的量的总和、MBR(最小外接矩形，minimum bounding rectangle)的长与宽、线段以连线为基础的对称性等，该对称性指标还能进一步说明该线的自相似性的范围与程度。因为地理现象的特殊性，即使是同一条线，自相似性有一定的存在范围，也即在某一给定的尺度范围内自相似性才会存在，那么在线综合时如何保持

线的分形特征就受到学者的关注。在计算线的长度时，不同步长，得到不同的长度，即不同精度条件下线的长度不同。通常线的长度随精度的降低而减少。

另外，曲线在进行多尺度化简时并没有发生自相交的现象，即拓扑几何特征没有变化，这也是在进行曲线综合时要保持的基本特征之一。

综上所述，线状要素多尺度表达的形状变化问题可以从它的长度、方向、曲直、自相似性、拓扑特性等方面进行分析。

4.2.3　线状要素多尺度表达位置不确定性的评价方法

线元位置不确定性的研究，始于 Chrisman 在 1982 年提出的“ε-带”指标，经过多年的发展，人们基于不同的数学工具逐步发展了“E-带”、“e-带”、“S-带”、“De-带”、“H-带”等理论。对于线元位置不确定性的误差带模型，现在发展得十分完善。各种曲线(直线、圆曲线、缓和曲线、一般拟合曲线等)的误差带模型均被建立；各种误差来源(数据、模型等)均被考虑进去；误差带宽度的确定运用了各种数学方法(熵理论、信息论等)。更有学者研究了误差带的形状、代数解析方法等。这些都是非常成熟的评价初始线的位置不确定性的方法，但是对于线状要素的多尺度表达的位置不确定性，现在相关研究成果主要集中在评定简化线相对于初始线的差异上。而事实上，除了化简，光滑也是实现线状要素多尺度表达的重要综合操作算子。本小节综合考虑初始线的位置不确定性、化简和光滑操作引入的不确定性，构建出线状要素多尺度表达综合不确定性的评价方法。

1. 简化线不确定性评价方法

线状要素多尺度表达是通过自动综合来实现的。线综合还包括化简、光滑、合并和移位(McMaster，1987b)。然而，化简是可论证的最常应用的线综合操作，被广泛应用于商业 GIS 软件包中。因而，在现有的线状要素多尺度表达不确定性评价方法中，主要是对线化简引起的几何不确定性的评价。McMaster(1986)描述一系列比较一条线简化过程前后几何特性的指标：长度变化百分比(简化线

总长度除以初始线总长度的百分比）；坐标数量变化的比率，每英寸坐标数量的平均数差异和每英寸坐标数量标准差变化比率（密度对比）；角度变化百分比（简化线连贯矢量间角度之和除以初始线角度之和）；曲线段变化比率（简化线曲线段数量除以初始线曲线段数量；曲线段是线上角度方向一致的部分，要么是正的，要么是负的）。这些指标被 Jasinski（1990）延伸：误差方差（初始线上的点到另一条线的平均垂线距离）；平均线段长度（线上点之间所有线段的平均长度）；平均棱角（范围从0到1）。这些方法是用线的长度、密度、角度、曲度来分析线的复杂度。还有一些通过对比来分析简化线和初始线之间位移的方法，如①矢量位移（McMaster，1987a，1987b；Jenks，1989）：化简过程中消除的初始线上的坐标点到简化线的垂线距离，该方法能提供沿线误差的统计分布信息，能得到直方图，并可计算误差的统计描述，如平均值、标准差。②变形多边形（McMaster，1987a，1987b；Buttenfield，1991）：原始线和简化线形成的多边形，相当于位移矢量方法采用间隔无限小的情形，显示位移矢量的无限密植距。矢量的平均长度等于变形多边形的总面积除以简化线的长度。这被称作"均衡距离失真"，因为它代表简化线均衡缓冲带的宽度，缓冲带的面积等于变形多边形的总面积。缓冲带的宽度是平均位移，即原始线和简化线之间的平均垂线距离。③关键距离（Little，1989）：原始线与简化线的距离大于特殊距离阈值的节点的比例。该指标是对超过特殊距离的节点比例的评估，而不是超过这个距离的程度。更重要的是，该方法假定在点上发展起来的误差度量方法也适用于线。它对每个节点的误差是独立评估的，尽管事实是所选节点是不相互独立的。变形多边形和位移矢量处理线是作为一个整体而不是一系列独立点的。

以上述这些度量方法和概念作为简化线位置不确定性评价的起源，到现在为止不同作者构建了若干简化线位置不确定性的评价方法，按时间顺序分别进行阐述，并进行分析比较。设初始线为OL，简化线为SL。

①Epsilon带方法（EBM）（Skidmore等，1992）：如图4.6（a）所示，用初始线到简化线的平均位移值作为epsilon带（Perkal，

1956)的宽度(ε)。平均位移值是简化线与初始线相交所形成的多边形的面积的和 (A) 与简化线长度(L) 的商(类似于面积位移(McMaster, 1986)), 如公式(4.3):

$$\varepsilon = \frac{\sum A}{L} \tag{4.3}$$

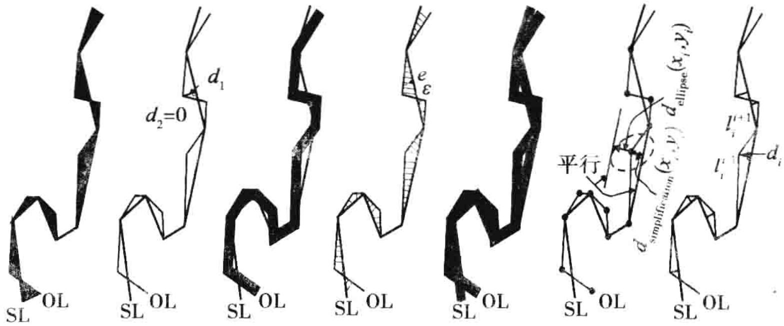


图 4.6 不确定性线和它的简化线的对比

这个方法基于 epsilon 带的一个估计给出一个平均位移值。

②Hausdorff 距离方法(HDM)(Abbas 等, 1995): 如图 4.6(b) 所示, 由一条线的节点到对应线的最大距离及其相反情况的最大距离构成。Hausdorff 距离是这些距离的最大值, 如公式(4.4):

$$D_{\text{DH}} = \max\ (\max(d_1);\ \max(d_2)) \tag{4.4}$$

从初始线上节点到简化线距离的平均距离(AD)是更容易获得的(McMaster, 1986; Mozas, 2008), 这个方法也可以用平均位移(ADs)代替最大位移。

③简单缓冲方法(SBM)(Goodchild, 1997)如图 4.6(c)所示, 围绕初始线(OL)生成一个缓冲区 B_{OL}, 并评价简化线在缓冲区内的长度, 如公式(4.5):

$$D_b = \%(\text{SL} \in B_{\text{OL}}) \tag{4.5}$$

通过增加缓冲区的宽度能获得一个简化线位于初始线缓冲区内的概率分布。

④中间点方法(PIM)(Tveite 等, 1999): 如图 4.6(d)所示，在简化线上每隔一个预先定义的距离(ε)生成一系列的中间点，计算每一个中间点(采样点)到初始线的距离，对于每一个样点，距离矢量 e 是这条线在这一点上的空间精度指标。

⑤缓冲区重叠统计法(BOS)(Tveite 等, 1999)如图 4.6(e)所示，又名双缓冲区方法(DBM)(Antonio T. Mozas 等, 2011)。该方法是围绕每一条线建立一个缓冲区，并且分析这些缓冲区相交时的所有可能情况。该方法中有几种方法如平均位移(公式(4.6))、振荡、完整性、错误编码都是从不断增加宽度的缓冲区中获得的。

$$AD = \pi b \frac{A(\overline{B_{SL}} \cap B_{OL})}{A(B_{SL})} \tag{4.6}$$

⑥最大距离方法(Chui Kwan Cheung 等, 2004): 如图 4.6(g)所示，该方法认为一条简化线的位置不确定性由两个因素所致：a. 处理过程中初始线位置不确定性的传播；b. 从初始线到简化线的偏差。计算 (x_i, y_i) 和 (μ_{x_i}, μ_{y_i}) 的偏差 $d_{\text{ellipse}}(x_i, y_i)$ 与初始线上符号点 (μ_{x_i}, μ_{y_i}) 到简化线 $SL_{j,j+1}$ 的最短距离 $d_{\text{simplification}}(x_i, y_i)$ 之和，简化线的不确定性指标 d_{position} 为

$$d_{\text{position}} = \max_i(d_{\text{ellipse}}(x_i, y_i) + d_{\text{simplification}}(x_i, y_i)) \tag{4.7}$$

在一般情况下，$d_{\text{simplification}}(x_i, y_i)$ 和 $d_{\text{ellipse}}(x_i, y_i)$ 的和是简化线上临近点到 (x_i, y_i) 的不确定性。因此，从初始线到简化线的最大偏差被定义为简化线上所有临近点到初始线上符号点的不确定性的最大值。这个不确定性指标评定了简化线的最大不确定性。

⑦节点影响方法(VIM)(Antonio T. Mozas 等, 2011): 如图 4.6(f)所示，该方法特别关注定义线的节点的重要性。它是基于初始线的节点到简化线的欧式距离。该距离用临近节点的线段的长度进行加权，这样得到的距离加权值更精确(公式(4.8))。对于所有的线，用线的长度计算加权算术平均值来得到这个评价位移值，又称平均 VIM 位移。

$$D = \frac{\sum_{i=1}^{n}(d_i(l_i^{i-1} + l_i^{i+1}))}{2L} \tag{4.8}$$

在这个方法中，加重了临近的线段长时节点的位移，缩小了临近节点线段短的位移。

综观这些方法，均是一个数量指标，分别从不同的角度评价了简化线与初始线的差异，除了最大距离法考虑了初始线的不确定性对简化线不确定性的影响外，其他方法均假定初始线是高精度的。

2. 线状要素多尺度表达综合不确定性评价方法

在空间数据库中，线基本上被定义为节点坐标串和线段，节点是定义线的结构元素。线状要素的多尺度表达是通过化简、光滑、合并、移位、符号化、可视化等综合操作算子来实现的。化简是对组成线状要素的坐标串进行处理，首先减少因缩小比例尺而变得冗余的数据，在此基础上删除或修改坐标串中次要点而保留主要特征点，这些被保留的点必须是制图资料上固有的；平滑算法可用来减少相邻值的差别；合并和移位操作是为了与其他目标相协调进行的综合操作；符号化和可视化是与相应比例尺和显示(物理)设备相适应的图形显示过程。由此可见，多尺度表达的线与初级尺度相比，其节点数量减少了，由于节点数量减少，节点之间光滑曲线段的位置与初始尺度的位置不相重合了。如图4.7所示，圆圈代表化简保留下来的节点，曲线为对简化保留节点进行光滑的结果。

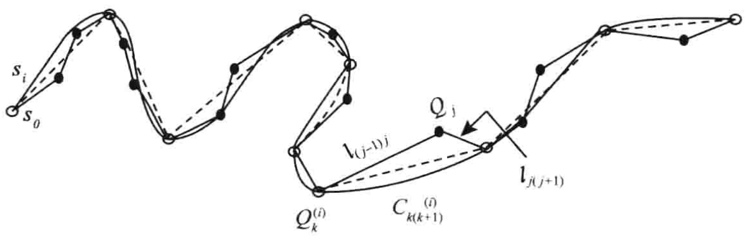


图4.7 初始线与简化尺度上曲线的对比

为了能很好地理解线状要素多尺度表达的位置误差，我们不考虑地图表示及物理设备对线状要素多尺度表达位置误差的影响，仅考虑抽象的数学意义上的线状要素多尺度表达位置误差的精度指标。根据李德仁(2004)“要区分用离散点逼近复杂曲线和曲面的逼

近误差与离散点自身的量测误差及其传播”的观点，对线状要素多尺度表达的不确定性评价时，一是要考虑节点的不确定性及其对整条线的影响(数据不确定性)；二是要考虑光滑曲线模型的不确定性。节点不确定性及其对整条线的影响相当于“离散点自身的量测误差及其传播”，光滑曲线模型的不确定性相当于“用离散点逼近复杂曲线的逼近误差”。因此，我们发展了一种适用于经过化简和光滑两个综合操作的线状要素(抽象的数学意义上的线)多尺度表达的位置不确定性评价模型。

(1)数据不确定性

初始尺度(s_0)上线的坐标节点一般由纸质地图数字化或野外实测(全站仪、动态GPS)而得到的。其余尺度($s_1 \sim s_m$)由初级尺度自动综合而来。若初始线上任意节点 Q_j ($1 \leqslant j \leqslant n_0$)的精度为 σ_{Q_j} ，初始尺度上由节点 Q_j 和 Q_{j+1} 构成的线段 $L_{j(j+1)}$ 上任意点的点位精度的计算参见各种误差带模型，如图4.7所示。

线状要素多尺度表达首先要进行的就是节点的化简，化简保留下来的节点，其位置不确定性与其对应的初始尺度上节点的不确定性相等。在尺度为 s_i 时，光滑曲线段 $C_{k(k+1)}(1 \leqslant k \leqslant n_i - 1)$ 上任意一点 R 的精度为该点法线方向(方位角为 θ)的中误差 σ_θ ，由光滑曲线段 $C_{k(k+1)}$ 两端点 $Q_k^{(i)}$ 、$Q_{k+1}^{(i)}$ 的位置误差传播而来，用一般拟合曲线的不确定性模型可计算出结果。虽然光滑曲线段上各点的数据不确定性从理论上讲是可以计算出来的，但是由于化简损失了线上若干节点，也就是数据信息的大量损失，再精确求解曲线上各点的数据不确定性意义不大。因此，根据光滑曲线端点对曲线上任意点的影响程度来定义光滑曲线段上任意点法线方向的不确定性。光滑曲线端点对曲线上点的影响程度与曲线上任意点到曲线端点的距离相关，点到端点的距离越近，受端点不确定性的影响越大，点到端点的距离越远，受端点不确定性的影响越小。因此，光滑曲线段 $C_{k(k+1)}^{(i)}(1 \leqslant k \leqslant n_i - 1)$ 上任意点 R 法线方向(方位角为 θ)的不确定性可定义为：

$$\sigma_\theta = \frac{r_k \sigma_{Q_{k+1}}^{(i)} + r_{k+1} \sigma_{Q_k}^{(i)}}{r_k + r_{k+1}} \tag{4.9}$$

上式中，r_k 为 R 点到 $Q_k^{(i)}$ 的距离，r_{k+1} 为 R 点到 $Q_{k+1}^{(i)}$ 的距离，如图4.8所示。

当曲线段 $C_{k(k+1)}^{(i)}$ 两端点的误差相等时，光滑曲线上各点的误差也相等，也就是 epsilon 带。

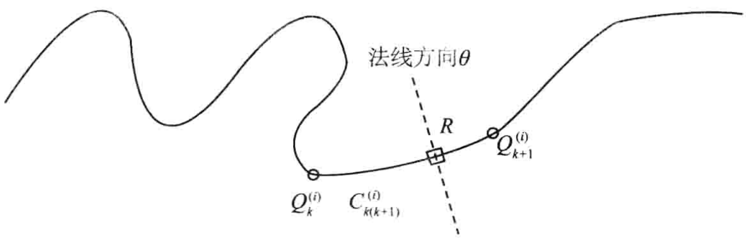


图4.8 光滑曲线上任意点及其法线方向示意图

按照式(4.9)，可以计算出光滑曲线段上各点的数据不确定性。但是在实践上，计算光滑曲线上无限数量点的误差带宽是非常困难的。而且由于化简损失了线上若干节点，也就是数据信息的大量损失，再单独求解曲线上各点的数据不确定性意义不大。因此，我们不再单独计算多尺度线上各点的不确定性，而是从整体出发，将多尺度线上所有节点不确定性的平均值作为多尺度线数据不确定性的评价指标。

多尺度线上的节点就是与其临近的两光滑曲线段的端点，临近光滑曲线段越长，说明该节点对整条线位置的影响越大。因此，这里用节点临近的曲线段长度对所有节点不确定性的平均值进行加权强化。节点不确定性的加权平均值见式(4.10)：

$$\sigma_i = \frac{\sum_{k=1}^{n_i} (\sigma_k(S_k^{k-1} + S_k^{k+1}))}{2S} \tag{4.10}$$

式(4.10)中，n_i 为尺度 s_i 上节点的个数，σ_k 为尺度 s_i 上第 k 个节点位置不确定性，S_k^{k-1} 和 S_k^{k+1} 分别为尺度 s_i 上与第 k 个节点相邻的两光滑曲线段的长度，S 为尺度 s_i 上曲线的总长度。在实际计算中，用相邻节点间的直线距离代替曲线段长度。

(2)模型不确定性

初始尺度上，线状图形的位置与真实位置是基本一致的。因此，可用简化尺度上曲线与初始尺度节点连线的位置偏差作为模型不确定性。这与蓝悦明、陶本藻(2003)定义二次曲线偏离端点连线的最大偏差为曲线光滑模型的不确定性，以及用简化线相对于初始线的某种位移作为简化线不确定性评价指标的思路是一致的。

计算机地图曲线中，光滑曲线一般为参数曲线，因此，简化尺度上曲线与初始线的距离为点到参数曲线的最短距离。逐点计算初始线上各点到简化尺度光滑曲线的最短距离，就构成了光滑曲线的模型不确定性。

逐点计算初始线上各点(节点及节点之外的各点)到简化尺度光滑曲线的最短距离从理论上讲是行得通的，但实现起来计算量非常大，而且数据之多不便于处理和理解。在线状要素多尺度表达中，最重要的是线位置变化对周围要素的位置和空间关系的影响。因此，也没有必要逐点计算模型不确定性。

我们将逐点计算模型误差的问题简化，仅用初始线上节点到简化尺度光滑曲线的最短距离的加权平均值作为模型误差(见式(4.11))，权重值为临近节点的线段的长度。这也与"节点是控制线的结构元素"的基本前提是一致的。

$$D_i = \frac{\sum_{j=1}^{n_0} \left(d_j \left(l_j^{j-1} + l_j^{j+1} \right) \right)}{2L} \tag{4.11}$$

式(4.11)中，n_0 为初始线上节点的个数，d_j 为初始线上第 j 个节点到尺度 s_i 上参数曲线的最短距离，l_j^{j-1} 和 l_j^{j+1} 分别为初始线上与第 j 个节点相邻的两线段的长度，L 为初始线的长度，如图 4.9 所示。

当简化尺度光滑曲线与初始线相交时，用上面的方法计算，一个线段将参与两个节点位移的计算。但是在交叉的情况下，平均位移是减少的。因此，在这种情况下，应该用两节点到交叉点的距离而不是线段的总长。

该方法增加了临近线段长的节点的位移，缩小了临近线段短的节点的位移。

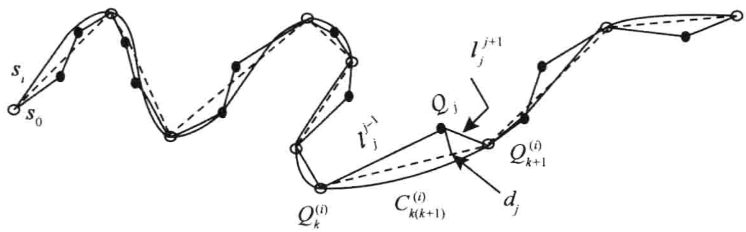


图 4.9　初始线上节点到简化尺度曲线的最短距离

(3)综合不确定性

尽管数据化简、曲线拟合是有先后顺序的，但是后一操作环节都不受上一环节的影响，也就是说，这些操作是相互独立的。因此，针对多尺度表达线的数据不确定性和模型不确定性，尺度 s_i 上线状要素的整体不确定性 $\sigma_{i-\text{whole}}$ 可以表示成一个基于误差传播定律的所有数据化简、曲线拟合的不确定性的函数，即

$$\sigma_{i-\text{whole}}^2 = \sigma_i^2 + D_i^2 \tag{4.12}$$

式(4.12)中，σ_i 为式(4.10)定义的尺度 s_i 上所有节点不确定性的加权平均值，D_i 为式(4.11)定义的初始线上所有节点到尺度 s_i 上光滑曲线的最短距离的加权平均值。

由于式(4.12)中 σ_i 和 D_i 的定义和计算与他们的理论值相比都进行了不同程度的简化。因此，尺度 s_i 上线状要素的整体不确定性 $\sigma_{i-\text{whole}}$ 也可以表达为简化曲线相对于初始曲线的整体平均位移 $D_{i-\text{whole}}$，即

$$D_{i-\text{whole}} = \sigma_i + D_i \tag{4.13}$$

综合不确定性指标 $\sigma_{i-\text{whole}}$ 和 $D_{i-\text{whole}}$ 综合了尺度 s_i 上数据的不确定性和模型的不确定性，而且在量化数据不确定性和模型不确定性时都加重了临近线段长的节点的位移，缩小了临近线段短的节点的位移。尺度扩展，数据量越来越少，量化其不确定性变得越来越困难，虽然这种不确定性的确定不是很严密，但这种线状要素多尺度表达不确定性的量化指标还是优于定性的判断。

在现有的研究中，有这样的共识：对于大比例尺地图，空间数

据精度高，不确定性带(即边界区)窄；而对于小比例尺地图，空间数据精度低，不确定性带宽。这无疑是对线状特征数据不确定性的定性描述，而没有明确不确定性带宽度的具体计算方法。不确定性带方法是评价线元不确定性的重要方法，尤其在进一步的空间分析和空间推理中有非常重要的作用，比如用不确定性带的边界作为空间目标粗集表达的上下近似边界，在空间目标粗集表达的基础上，可以进行空间方向不确定性表达及空间方向推理等。因此，可用各尺度曲线的综合不确定性大小作为相应尺度曲线不确定带的宽度，来解决空间目标多尺度表达不确定性空间分析和推理问题。

(4)实验与讨论

本文用来自比例尺为 1∶5000 的数字线划图进行实验，如图 4.10 所示，该线划图由 85 个坐标对构成。假定各节点的点位中误差均为 4.0m($\sigma_{Q_j} = 4.0\text{m}$，$1 \leqslant j \leqslant 85$)。

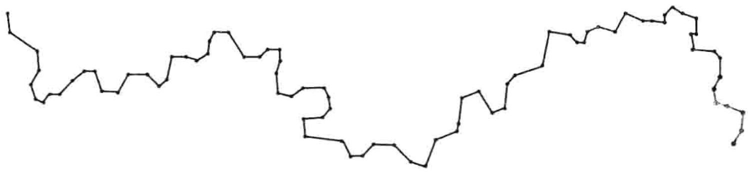


图 4.10　1∶5000 的数字线划图

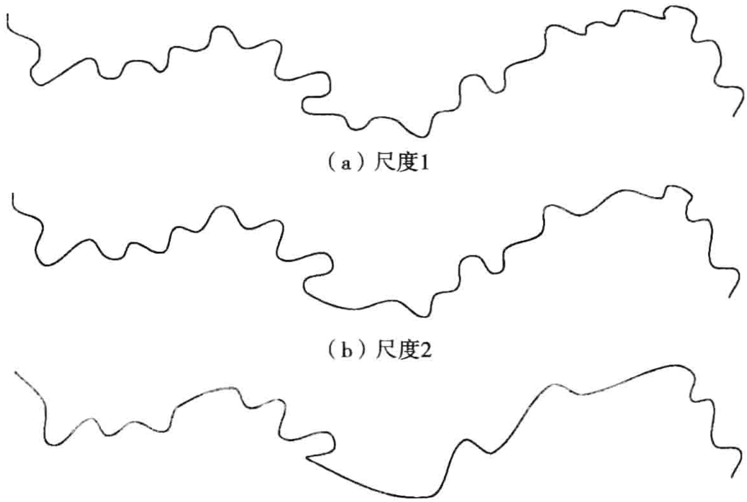
用 Douglas-Peucker 算法对原始 1∶5000 的数字线划图进行 3 个不同程度的化简，并分别用三次参数曲线进行光滑，结果显示如图 4.11，图 4.11(a)、(b)、(c)的特征点个数分别为 59、43、27。

利用式(4.10)进行计算得到图 4.11(a)、(b)、(c)的数据不确定性均为 4.0m。

利用式(4.11)进行计算得到图 4.11(a)、(b)、(c)的模型不确定性分别为 2.0m、6.5m、16.9m。

利用式(4.12)进行计算得到图 4.11(a)、(b)、(c)的整体不确定性分别为 4.47m、7.63m、17.37m。

利用式(4.13)进行计算得到图 4.11(a)、(b)、(c)的整体平



（a）尺度1

（b）尺度2

（c）尺度3

图 4.11 原始曲线的多尺度表达

均位移分别为 5.99m、10.5m、20.9m。

对本实验数据计算 Chui kwan Cheung 等人（2004）定义的最大距离 $d_{position}$ 、Antonio T. Mozas 等人（2011）定义的平均 VIM 位移 D ，其结果见表 4.1。

表 4.1 **不同方法计算的不确定性值**

不确定性		原图	尺度 1	尺度 2	尺度 3
综合不确定性	$\sigma_{i-whole}$/m	4.0	4.47	7.63	17.37
	$D_{i-whole}$/m	4.0	5.99	10.5	20.9
平均 VIM 位移（D/m）			2.0	6.5	16.9
最大距离法（ $d_{position}$/m ）		4.0	24.08	43.09	74.12

从表 4.1 可以看出，本文定义的综合不确定性指标值明显小于最大距离法的值，而大于平均 VIM 位移。平均 VIM 位移没有顾及

初始线的不确定性，而最大距离法考虑了初始线不确定性的传播，更重要的是该方法在考虑多尺度线相对于初始线的位移时只考虑了最大位移。而本书定义的指标一方面顾及了数据的不确定性（初始线不确定性传播至多尺度曲线的不确定性），另一方面数据不确定性是多尺度表达的线相对于初始线的平均位移，平均位移比最大位移小。

4.2.4　线状要素多尺度表达形状不确定性的评价方法

目前，已有若干描述线形状特征的方法，这些方法可以分为两大类，一类是线状要素复杂度描述方法（王桥等，1998；郭仁忠，2001），另一类是曲线长度归算方法（郭仁忠，2001；王桥，1996）。这些方法仍然可用于描述多尺度表达的线的形状特征。曲线尺度变化时，这些方法对应的指标值的变化量就是线状要素多尺度表达形状不确定性的评价指标。

1. 线状要素复杂度描述方法

线状要素是地图图形要素的基本要素，如海岸线、河流、等高线、境界线等都是地图图形数据处理中经常遇到的对象。为了对它们进行分析和处理，人们一直设法来描述它们的形状结构特征，目前在地图图形数据处理中常用的方法有：

①定性描述法：即以语言文字对曲线的形状进行描述。例如，用复杂、规则、平滑等描述曲线的复杂程度，用蛇行、菌形、套形、弧形等描述曲线的形状。

②样图法：对曲线按复杂程度进行编号，并制成样图来作为曲线复杂程度的测量标准。Chernyaeva（1958）给出了 7 种不同复杂程度的标准，如图 4.12 所示，人们可凭目估来对照所研究的曲线，从而采取不同的处理。

③基本弯曲系数法：设 l_i 表示曲线第 i 个基本弯曲（曲线矢长为 0.2mm 以上者）的个数，p_i 表示第 i 个基本弯曲的弦长（$i = 1, 2, \cdots, n$），则定义曲线基本弯曲系数为：

$$K = \sum_{i=1}^{n} l_i \Big/ \sum_{i=1}^{n} p_i$$

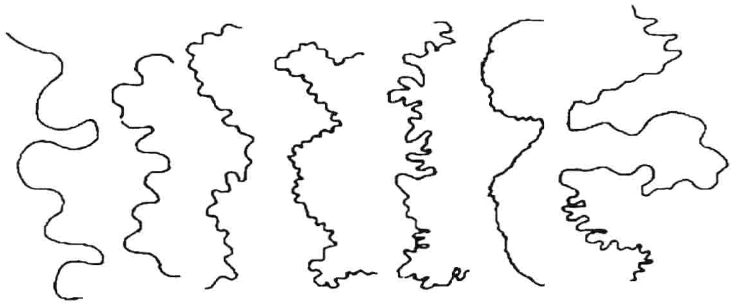


图 4.12 曲线样图举例

④弯曲个数法：设单位长度 L 上的基本弯曲个数为 n，则定义曲线的弯曲程度为：

$$S = n/L$$

⑤曲率度量法：线状物体的曲率由数学分析定义为曲线切线方向角相对于弧长的转动率。设曲线的形式为 $y = f(x)$，则曲线上任意一点的曲率为：

$$K = \frac{y''}{(1 + y'^2)^{2/3}} \tag{4.14}$$

对于以参数形式 $x = x(t)$，$y = y(t)(\alpha \leqslant t \leqslant \beta)$ 表示的曲线，其上任意一点的曲率的计算式为：

$$K = \frac{x'y - xy'}{(x'^2 + y'^2)^{3/2}} \tag{4.15}$$

根据公式(4.14)和(4.15)计算曲线曲率的前提是曲线必须是连续光滑的(二阶导数存在)，对于多项式曲线，显然必须是二次以上的光滑曲线。

对于以离散点表示的线状物体，必须先进行光滑插值，然后方可按照公式(4.14)和(4.15)计算。曲率反映了曲线的局部弯曲特征，为了反映曲线的整体弯曲特征，应该计算曲线的平均曲率。根据对弧长的曲线积分的定义，光滑曲线 L 的平均曲率可按式(4.16)计算：

$$\bar{K}=\frac{\int_L K\mathrm{d}s}{\int_L \mathrm{d}s} \tag{4.16}$$

对于以参数形式表示的曲线，按式(4.16)的转换形式计算非常复杂，简单方法是将$[\alpha,\beta]$ n 等分，得到

$$\alpha=t_0,\ t_1,\ \cdots,\ t_n=\beta$$

进而根据(4.14)式，在各分段上选取某一点计算曲率，得到K_0、K_1、…，K_n，则

$$\bar{K}=\frac{I}{N}\sum_{i=1}^{n} k_i$$

另外，设曲线长度为 L，相应的平稳曲线(通过拐点所作的内插的平滑曲线)的长度为 L'，曲线的闭合直线为 d，如图 4.13 所示，定义曲线的相对曲率为：

$$w_1=L/L'$$

曲线的总轮廓曲率为：

$$w_2=L'/d$$

曲线的总曲率为：

$$w_3=L/d$$

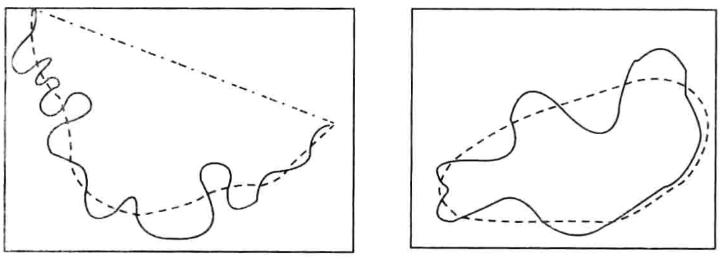


图 4.13　平滑曲线与闭合直线

⑥分维数法：作为地学图形的线状要素广泛存在着自相似性(统计自相似性)，适用于用分形分析方法对其进行研究。可以用连续取值的分维数来描述不同复杂程度和不同形状特征的曲线。设

所考虑分形曲线的长度为 L，分维数为 D，用码尺 S 去度量该曲线，可得到

$$\log L = (1 - D)\log S + \log K$$

只要对码尺的一个单调递增序列 $\{S_i\}_{i=1}^{M}$，再用线性回归模型拟合点对 $(\log S_i,\ \log L(S_i))(i = 1,\ 2,\ \cdots,\ M)$，即可求出 D 值。

⑦曲折率法：首先约定描述数字化曲线的点列 $\{A \mid A_k:\ k = 1,\ 2,\ \cdots,\ n\}$ 等价于由折线 A_1A_2，A_1A_3，…，$A_{n-1}A_n$ 构成的曲线。定义单位长度曲线所转过的总角度为曲线的曲折率 ξ：

$$\xi = \frac{\lambda \sum_{k=2}^{n-1} \alpha_k}{\sum_{k=1}^{n-1} |\overline{\gamma_k}|}$$

其中，λ 为比例尺变比（$\lambda \geqslant 1$），α_k 和 $|\gamma_k|$ 为折线 A_kA_{k+1} 的转角和长度。

$$\alpha_k = \arccos\left(\frac{\bar{\gamma}_k + \bar{\gamma}_{k+1}}{|\bar{\gamma}_k||\bar{\gamma}_{k+1}|}\right),\quad k = 2,\ 3,\ \cdots,\ n - 1$$

其中，$\bar{\gamma}_k = (x_{k+1} - x_k)\bar{i} + (y_{k+1} - y_k)\bar{j}$，$|\bar{\gamma}_k| = [(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2]^{\frac{1}{2}}$，$\bar{i} = (1,\ 0)$ 和 $\bar{j} = (0,\ 1)$ 为单位坐标矢量，$\bar{\gamma}_k + \bar{\gamma}_{k+1} = (x_{k+1} - x_k)(x_{k+2} - x_{k+1}) + (y_{k+1} - y_k)(y_{k+2} - y_{k+1})$ 为标量积。

⑧曲折差法：首先定义曲线的取中位线算子 ∇_{m}。对于由点列 $\{A(n) \mid A_k:\ k = 1,\ 2,\ \cdots,\ n\}$ 描述的曲线，将 A_k，A_{k+1} 两点连线的中点记为 $A_k^1 = \mathrm{med}_k\{A_k,\ A_{k+1}\}$。则由点列 $\{A^1(u) \mid A_k^1:\ k = 1,\ 2,\ \cdots,\ u\}$ 描述的曲线 $A^1(u)$ 称为原曲线 $A(n)$ 的一阶中位线，并记为 $A^1(u) = \nabla_{\mathrm{m}}A(n)$。类似地，$d$ 阶中位线的算子表达式为 $A^d(u) = \nabla_{\mathrm{m}}^d A(n)$。显然，$d$ 阶中位线算子能满足 $\nabla_{\mathrm{m}}^d A(u) = \nabla_{\mathrm{m}}^{d-1}(\nabla_{\mathrm{m}}A(n))$，这里规定 $\nabla_{\mathrm{m}}^0 = 1$。若 $\{A(n) \mid A_k:\ k = 1,\ 2,\ \cdots,\ n\}$ 为具有 n 个点的任一非封闭凸性曲线，则 $A(n)$ 最多经过 $d = (n - 2)$ 阶 ∇_{m}^d 运算后必为直线，不妨称该直线为标准中位线。

显然，曲线"偏离"其标准中位线的程度可以反映曲折的大小。在这里，"偏离程度"一词可以有不同的理解。例如，可取原曲线上各点到标准中位线的偏差绝对值之和，或距离之和，或其中的最

大距离等衡量。这里，可借用方差概念定义曲折差来度量曲折的大小：

$$\sigma_{\xi} = \frac{1}{\lambda}\sqrt{\frac{\sum_{k-1}^{n} e_k^2}{n}}$$

其中，e_k 为原曲线 $A(n)$ 上第 k 点 A_k 到标准中位线的距离。设标准中位线的端点为（$\bar{x}_1$，$\bar{y}_1$）和（$\bar{x}_2$，$\bar{y}_2$），则其方程为：

$$\begin{vmatrix} x & y & 1 \\ \bar{x}_1 & \bar{y}_1 & 1 \\ \bar{x}_2 & \bar{y}_2 & 1 \end{vmatrix} = 0$$

而一般式为 $(\bar{y}_1 - \bar{y}_2)x + (\bar{x}_2 - \bar{x}_1)y + (\bar{x}_1\bar{y}_2 - \bar{y}_1\bar{x}_2) = 0$，于是有：

$$e_k = \frac{|(\bar{y}_1 - \bar{y}_2)x_k + (\bar{x}_2 - \bar{x}_1)y_k + (\bar{x}_1\bar{y}_2 - \bar{y}_1\bar{x}_2)|}{\sqrt{(\bar{y}_1 - \bar{y}_2)^2 + (\bar{x}_2 - \bar{x}_1)^2}} \tag{4.17}$$

式(4.17)中，$k = 1, 2, \cdots, n$。实际应用中，也可用一阶中位线代替标准中位线计算 e_k，而且许多规则曲线的一阶中位线本身就是直线。但应当注意，在一阶中位线是非直线时，它本身也存在着复杂性问题。此外，有时我们关心单位长度曲线总曲折的大小，这可用相对曲折差衡量，即

$$\eta = \frac{\sigma_{\xi}}{\frac{1}{n-1}\sum_{k=1}^{n-1}|\bar{\gamma}_k|}$$

⑨复杂度法：曲折率和曲折差分别从曲折多少和曲折大小两个方面衡量曲线的复杂度。综合曲折率和曲折差，定义复杂度来衡量曲线的整体复杂性，即

$$\omega = \sigma_{\xi}\sum_{k=1}^{n-1}\alpha_k$$

此外，还有 Fourier 描述子，这里不再作介绍。

上面是对现有的部分曲线形状(复杂度)描述的指标的简单介

绍，实际上它们都有不合理的因素存在。曲率、基本弯曲系数、弯曲个数是在平均意义上的，过于粗糙，很难反映曲线形状结构的本质特征。Fourier 描述子将时空域的地图曲线转换到频率域，用周期函数的 Fourier 级数展开式量化曲线的形状特征，优点是能描述曲线的细部复杂性，特别适用于高复杂性曲线，缺点是计算困难。分维数适用于度量不可微自相似曲线的复杂程度，而且对于某些非分维数图形，如果把尺度范围限制在特征长度以内，也可能找到形式上类同于分维数的“拟分维数”，优点是能描述地图上许多自然边界的复杂性，如河流、海岸线等，缺点是分维数的各种求法工作量较大，如粗视化法、测度关系法、相关函数法等。曲折率和曲折差分别从曲折多少和曲折大小两个方面衡量曲线的复杂度，均与尺度有关，且相互制约、互相补充。复杂度是对曲折率和曲折差的综合，可以用来衡量地图曲线的整体复杂性。

2. 曲线长度归算方法

线状物体的长度是最基本的形态参数之一，在矢量数据下，线状物体 L 表示为坐标串 (x_i, y_i) 或者是 (x_i, y_i, z_i)，$i = 1, 2, \cdots, n$，则线状地物长度计算公式为(不考虑比例尺)：

$$L = \sum_{i=0}^{n-1} \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2 \right]^{1/2} = \sum_{i=1}^{n} l_i$$

在大多数情况下，长度计算限制在二维空间中，用式(4.18)计算长度的更多。

$$L = \sum_{i=0}^{n-1} \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right]^{1/2} = \sum_{i=1}^{n} l_i \quad (4.18)$$

在地图图形数据处理中，往往需要把一种比例尺下的曲线长度归算到另一种比例尺或实地的长度，这对于空间分析、地图应用等都具有重要的意义。然而，由于制图综合的影响，使这一问题变得比较复杂。自 20 世纪 50 年代以来，就不断有学者对它进行探讨，但是，在计算机制图与地理信息自动化技术迅速发展的今天，从合理性、实用性等诸多方面看，现有方法都存在一些明显的局限，其主要表现在两个方面：一是基于人工制图，二是较难顾及曲线本身的形状特征。例如，指定比例尺曲线长度归算法需要根据标准曲线

样图，目估确定归算系数，这显然带有一定的主观性，且不适于计算机自动处理；伏尔科夫归算长度公式与利用三种比例尺归算长度公式，需要用到多种成等比关系的比例尺系列地图数据，不但不够方便，而且不能够充分顾及河流的不同类型、河段；沃洛比也夫方法虽然考虑了随比例尺增长的岸线长度增长的百分比 n(即在一定程度上顾及了曲线自身的特点)，但 n 的确定不具有普遍性；比例倍数法也存在类似的问题，该方法的归算系数是基于对公路要素的研究得出的，不一定能够很好地适合其他要素及不同复杂程度的曲线。基于分维估值的曲线长度归算方法是顾及曲线本身形状结构特征的长度归算公式，下面对其原理作简单介绍。

正如本章 4.2.4 节所指出的，地图曲线具有分形性质，即在不同尺度上都具有细节，可以用分维对其形状特征进行量化。根据分形理论的基本原理，有

$$L(\gamma)=C\gamma^{1-D},\ \gamma\in[\gamma_1,\ \gamma_2] \tag{4.19}$$

其中，$L(\gamma)$ 是以 γ 为尺度所量测的曲线的长度；D 即为曲线的分维；C 是常数；$[\gamma_1,\ \gamma_2]$ 是无标度区。

一般地说，当以 $\gamma\in[\gamma_1,\ \gamma_2]$ 为尺度量测曲线时，所得的长度值是不确定的，即此时没有特征长度存在；而用小于无标度区间下限 γ_1 的尺度去量测曲线时，则可得到较稳定的长度值——极限长度。实用中，往往取具有实际意义的充分小的 $\gamma_0<\gamma_1$ 为量测尺度，以此来获得曲线极限长度。

设在比例尺为 $1:M_1$ 和 $1:M_2$ 的地图上，用 γ_0 量测所研究的曲线，得到的长度分别为 L_{M_1} 和 L_{M_2}，易知，γ_0 的实际尺度分别为 $d_1=\gamma_0 M_1$ 和 $d_2=\gamma_0 M_2$，一般地，有 $d_1,\ d_2\in[\gamma_1,\ \gamma_2]$。从而，由式(4.19)，可得

$$L_{M_1}=C(M_1\gamma_0)^{1-D} \tag{4.20}$$

$$L_{M_2}=C(M_2\gamma_0)^{1-D} \tag{4.21}$$

式(4.20)与式(4.21)相比，可得

$$L_{M_1}/L_{M_2}=(M_1/M_2)^{1-D}$$

或

$$L_{M_1}/L_{M_2}=(M_2/M_1)^{D-1} \tag{4.22}$$

式(4.22)即给出了一种基于分维的不同比例尺地图上曲线长度变化的规律，显然，它可用于长度归算。对于给定的 M_1、M_2 及 L_{M_2}，当利用本章2.4.1节所述的方法求出曲线分维值 D 后，即可由式(4.22)确定或归算出 L_{M_1}。当然，式(4.22)是在综合前后分维值保持不变(在统计意义上)的前提下使用的，在实际应用中，常见的比例尺跨度范围内是可以满足式(4.22)的使用前提的。

由于 D 从数量上刻画了曲线的形状结构特征或复杂程度，所以，用式(4.22)确定的曲线长度归算结果，不仅与 M_1、M_2 及 L_{M_2} 有关，而且，还与反映曲线本身形状结构特征的数量指标有关，这正是公式(4.22)的主要优点。事实上，比例倍数法就是式(4.22)中 $D = 1.017$ 的特例。此时，D 即为道路的分维值。而实际上，不同地图曲线的分维值可以是相差很大的，如沉降型海岸线，分维值 D 可达1.5之多。因此，式(4.22)能够充分顾及不同地图曲线自身的形状特征，具有更为普遍的意义。

4.3 面状要素多尺度表达的不确定性

地球表面的许多物质是以连续分布的形式覆盖在大地上的，被称为面状地理现象或面状要素，地图上用面状符号表示。用面状符号表示的有水部范围、林地范围、土地利用分类范围、各种区划范围、动植物和矿藏资源分布范围等。

面状要素在地图上是由线段构成的封闭多面形的边界形成的。面状要素的位置不确定性由其多边形边界的位置误差所决定。已有研究用两种方法来讨论多边形的位置不确定性，一种方法是考虑具有位置误差的多边形的面积；另一种方法则是考虑多边形误差带模型。因此，面状要素多尺度表达的不确定性也是从两个方面进行研究的，一是多边形多尺度表达的面积不确定性，二是多边形多尺度表达的误差带模型。由于多边形的误差带模型是线元误差带模型的扩展，因此，多边形多尺度表达不确定带模型是线状要素多尺度表达不确定带模型的扩展。在4.2.3节建立的线状要素多尺度表达位置不确定性的评价方法中，可直接用于多尺度面的不确定性评价。

因此，本节仅研究多边形多尺度表达的面积不确定性。

4.3.1 多边形的面积精度

设多边形 A 具有 n 个顶点 $P_i=(x_i,\ y_i)$ 和 n 条边 P_1P_2，P_2P_3，…，$P_{n-1}P_n$ 以及 P_nP_1，单个顶点的方差-协方差矩阵具有如下形式：

$$\sum_{P(i)P(i)}=\begin{bmatrix}\sigma_{x(i)}^2 & \sigma_{x(i)y(i)}\\ \sigma_{x(i)y(i)} & \sigma_{y(i)}^2\end{bmatrix}$$

多边形面积 S 为：

$$S=\frac{\sum_{i=1}^{n}[x_i(y_{i+1}-y_{i-1})]}{2}=\frac{\sum_{i=1}^{n}[x_i\Delta y_{i-1,\ i+1}]}{2}$$

面积的微分为：

$$\mathrm{d}S=\frac{\sum_{i=1}^{n}[\Delta y_{i-1,\ i+1}\mathrm{d}x_i+\Delta x_{i-1,\ i+1}\mathrm{d}y_i]}{2}$$

多边形面积的方差：

$$\sigma_{S1}^2=\frac{\sum_{i=1}^{n}[\Delta y_{i-1,\ i+1}^2\sigma_{x(i)}^2+\Delta x_{i-1,\ i+1}^2\sigma_{y(i)}^2+2\Delta x_{i-1,\ i+1}\Delta y_{i-1,\ i+1}\sigma_{x(i)y(i)}]}{4}$$

当 $\sigma_{x(i)}^2=\sigma_{y(i)}^2=\sigma_0^2$ 且 $\sigma_{x(i)y(i)}=0$ 时，该方差式可简化为：

$$\sigma_{S1}^2=\frac{\sum_{i=1}^{n}[\Delta y_{i-1,\ i+1}^2+\Delta x_{i-1,\ i+1}^2]\sigma_0^2}{4}=\frac{\sum_{i=1}^{n}l_{i-1,\ i+1}^2\sigma_0^2}{4}\tag{4.23}$$

其中，$l_{i-1,\ i+1}$ 是顶点 P_{i-1} 和 P_{i+1} 之间的距离。对于所有 $i=1$，2，…，n，n 边规则多边形边长 $l_{i-1,\ i+1}$ 的计算方法相同，且边长值相等，设为 l。于是，规则多边形面积的标准差为：

$$\sigma_{S1}=\sqrt{n}\sin\left[\frac{\pi}{2}-\frac{\pi}{n}\right]\cdot l\cdot\sigma_0\tag{4.24}$$

从方程(4.24)可以看出，多边形面积是多边形边长的函数。这表明多边形面积的不确定性可由这些边长的位置误差所决定。下面将阐述多边形面域位置不确定性研究中边长位置误差对其的影响。

令 Δ_i 表示多边形 A 中 $P_i P_{i+1}$，$i = 1, 2, \cdots, n$ 的误差带面积，多边形面域的位置误差 Δ_S 与边 $P_i P_{i+1}$ 误差带面积 Δ_i 之间的关系如下式：

$$\Delta_S = \Delta_1 + \Delta_2 + \cdots + \Delta_n$$

若 Δ_1，Δ_2，…，Δ_n 相互独立，则多边形面积的标准差为：

$$\sigma_{S1} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2} \tag{4.25}$$

其中，σ_i 为边 $P_i P_{i+1}$ 对应误差带面积的标准差，σ_i 的值为 2.2.1 节中导出的标准误差带面积的一半值，即 $\sigma_i = \dfrac{\Delta_i}{2}$。对于具有周长 $4l$ 的 n 边规则多边形而言，多边形面积的标准差为：

$$\sigma_{S2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2} = \frac{\sqrt{n}\,\Delta_S}{2n} = \frac{\Delta_S}{2\sqrt{n}} \tag{4.26}$$

式(4.25)和式(4.26)表明，多边形面域的位置不确定性随着多边形边长位置误差的增大而增大。

4.3.2 边界化简后多边形的面积精度

多边形的多尺度表达也是分两步完成的，首先是多边形边界线的化简，其次是光滑。为了简单起见，这里不考虑边界光滑问题。

在 4.2.3 节的内容中构建了线状要素化简后位置不确定性的计算方法，多边形化简后，可用同样的方法计算其位置不确定性，然后将其代入式(4.23)或者式(4.25)计算简化后多边形的位置不确定性。

设多边形 A 具有 n 个顶点 $P_i = (x_i, y_i)$ 和 n 条边 $P_1 P_2$，$P_2 P_3$，…，$P_{n-1} P_n$ 以及 $P_n P_1$，单个顶点的方差-协方差矩阵具有如下形式：

$$\sum_{P(i)P(i)} = \begin{bmatrix} \sigma_{x(i)}^2 & \sigma_{x(i)y(i)} \\ \sigma_{x(i)y(i)} & \sigma_{y(i)}^2 \end{bmatrix}$$

多边形 A 化简后形成的 A' 有 m 个顶点 $P_j = (x_j, y_j)$ 和 m 条边 $P_1 P_2$，$P_2 P_3$，…，$P_{m-1} P_m$ 以及 $P_m P_1$，按照 2.3 节的方法，化简后多边形边界上每个点的不确定性是相同的，均为 σ_{whole}，于是多

边形 A 化简后的面积方差为：

$$\sigma_{S1}^2=\frac{\sum_{j=1}^{m}[\Delta y_{j-1,\ j+1}^2+\Delta x_{j-1,\ j+1}^2]\sigma_{\text{whole}}^2}{4}=\frac{\sum_{j=1}^{m}l_{j-1,\ j+1}^2\sigma_{\text{whole}}^2}{4}$$

其中，$l_{j-1,\ j+1}$ 是顶点 P_{j-1} 和 P_{j+1} 之间的距离，σ_{whole} 为简化后多边形节点不确定性 σ 与简化多边形相对于初始多边形的平均距离 D 合成，即

$$\sigma_{\text{whole}}^2=\sigma^2+D^2 \tag{4.27}$$

公式(4.27)中，$\sigma=\frac{\sum_{j=1}^{m}(\sigma_j(l_j^{j-1}+l_j^{j+1}))}{2L_{A'}}$，其中 m 为简化多边形 A' 上节点的个数，σ_j 为简化多边形 A' 上第 j 个节点位置不确定性，l_j^{j-1} 和 l_j^{j+1} 分别为简化多边形 A' 上与第 j 个节点相邻的两边长的长度，$L_{A'}$ 为简化多边形 A' 的总边长。公式(4.27)中，

$$D=\frac{\sum_{i=1}^{n}(d_i(l_i^{i-1}+l_i^{i+1}))}{2L_A}$$

其中，n 为初始多边形 A 节点的个数，d_i 为初始多边形 A 上第 i 个节点到简化多边形 A' 的最短距离，l_i^{i-1} 和 l_i^{i+1} 分别为初始多边形 A 上与第 i 个节点相邻的两线段的长度，L_A 为初始多边形 A 的总边长。

4.4　点群多尺度表达的不确定性

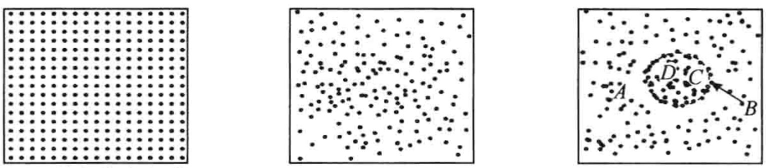
在地图中，空间的山包群、风蚀残丘群、成群分布的岛屿、湖泊、沙丘群、分散或散列的建筑物，以及专题图斑等，整体上都呈点状空间分布。随着人类认知和需求层次的日益提高，对点群的研究也延展到了多个尺度上，即点群的多尺度表达。

4.4.1　点群及其空间分布特征

“点群”中点的个体特征和属性不是点群研究关心的内容，而是关注于点群所体现出来的群体特征，即针对点群进行的空间分布

分析，包括空间分布格局和空间分布结构化信息。分析的目的在于提取空间相关规律，如根据病人的分布寻求分布中心表达的病源，进而在空间上寻求与水源、矿物成分等自然条件的相关特征。根据空间域内包容分布对象的能力，顾及相互间的影响作用确定最佳设施配置（如城市规划中对医院、菜场、学校等公共设施的用地选址即属此情形）。

点群的空间分布格局通常可以分为均匀分布、随机分布和集聚分布三种类型，如图 4.14 所示，可以用样方内点数均值变差、点间最近距离均值、点密度距离函数，或点密度距离分形维数来度量，Ripley's K 函数可以分析任意尺度的空间分布格局，成为分析点要素分布格局最常用的方法（Ripley，1981）。Ripley's $K(d)$ 定义为



（a）均匀分布点群　（b）随机分布点群　（c）集聚分布点群

图 4.14　点群空间分布类型

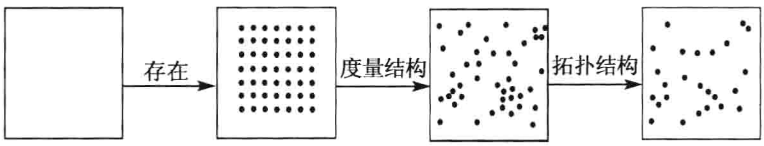
$$K(d) = (\text{在距离为 } d \text{ 的范围内事件平均数目})/\lambda = A\sum_{i}^{n}\sum_{j}^{n}\frac{w_{ij}(d)}{n^2} \tag{4.28}$$

其中，λ 为事件发生强度；n 为点的个数；d 为距离；$w_{ij}(d)$ 为空间事物个体 i 与个体 j 之间的距离；A 为研究区域面积（Ripley，1977）。公式（4.28）中的核心变量 $w_{ij}(d)$ 是由点群中所有点的位置数据计算出来的。

描述点群空间分布结构化信息的特征参量包括分布范围、数量、密度及其分布、分布中心、分布轴线、子群数量、子群的层次性结构等，Zhilin Li 还提出用几何信息量来描述点群空间分布结构化的信息量。崔世林提出了一种基于滑动窗口的局部分维分析方

法——元分维模型来分析点群的空间分布特征。

艾廷华等人(2002)从信息论的观点出发把群点分布隐含的信息内容划分为存在信息、度量结构信息和拓扑结构信息 3 种。如图 4.15 所示，由 A→B 得到“区域内具有 n 个点”，此时不考虑点如何分布，获得的是存在性信息；B→C 得到点分布以距离远近表达的空间结构，获得的是度量结构信息；由 C→D 得到相邻点之间的性质差异性，与距离方向无关，获得点分布的拓扑结构信息。



A　　B　　C　　D

图 4.15　点群分布隐含的 3 种信息内容

4.4.2　点群多尺度表达方法

点群多尺度表达的主要目的仍然是进行点群数据空间分布分析，也就是在不同的观察尺度上进行点群数据分布分析，为空间决策提供信息。点群的多尺度表达通过化简操作来实现，一是选取数量的问题，用著名的开方根规律来解决；二是“选取哪些”的问题，即具体保留哪些点、删除哪些点，属于度量结构信息的化简问题。在群点化简综合中，从大比例尺到小比例尺，存在性信息必然减少(不确定熵增大)，而度量结构信息和拓扑结构信息要最大限度地保持，达到用较少点表达原群点描述的分布疏密、分布范围、分布中心等结构信息的目的。在居民地群点化简中，有重力模型法(gravity modeling)、分布相关系数控制法(distribution-coefficient-control)，圆增长法(circle growth)。这些方法的基本思想是用圆表示点的影响区域，通过圆之间的关系考察群点之间影响作用域进而判断化简中哪些点保留、哪些点删除。毋河海(1997)运用凸壳算法建立了群点的圈层结构，将群点的化简分解为各凸壳圈上点的化

简问题。更简单的方法则是对群点区域进行网格划分，运用随机函数对各网格内的点重采样。艾廷华(2002)利用 Delaunay 及其对偶 Voronoi 图定义群点分布的范围、密度、分布中心、分布轴线及其计算方法，并建立了 Voronoi 图动态重建进行群点化简的方法。

4.4.3 点群多尺度表达不确定性的评估

点群数据本身的不确定性包括位置不确定性和属性不确定性。查阅各特征参量的计算方法，不难得出，反映点群空间分布的特征参量都是针对点群的位置数据所作的统计分析。故点群空间分析结论的不确定性为其位置不确定性在计算过程中传播的结果。依据点群的位置不确定性及其传播机理，就能得出点群空间分析结论的不确定性。图 4.16 反映了针对点群进行的多尺度表达、空间分析以及不确定性分析流程。

文中点群多尺度表达不确定性研究的“点群”对应图 4.16 中“抽象的点群数据”和“多尺度(比例尺)的点群数据”。“抽象的点群数据”不确定性研究的内容为位置不确定性和空间分析结论不确定性，“多尺度(比例尺)的点群数据”不确定性研究的内容为位置不确定性和空间分析结论不确定性。通过对不同尺度点群位置不确定性和空间分析结论不确定性的评估，可以得出点群位置及其空间分析结论的不确定性随尺度变化的规律。

1. 点群综合质量评估方法

对点群目标的综合质量评估，有许多指标，如最小形状变形、最小位移、图形覆盖密度、综合前后目标个数比、综合前后目标之间的平均距离、频率、平均数、数学期望、方差/标准差、变差系数的差异等。钱海忠(2005)提出一种基于极化变换的点群几何质量评估方法，在点群特征点保持、整体性保持两个方面提出了量化的质量评估指标，以及指标的量化方式，用特征点保持率、聚类数量保持率、聚类轮廓保持率、点目标删除率作为点群自动综合质量的评价指标，用于进一步的质量评估。

2. 点群多尺度表达位置不确定性的评估

一个点是空间信息的最简单形式之一，在地理空间中往往需要

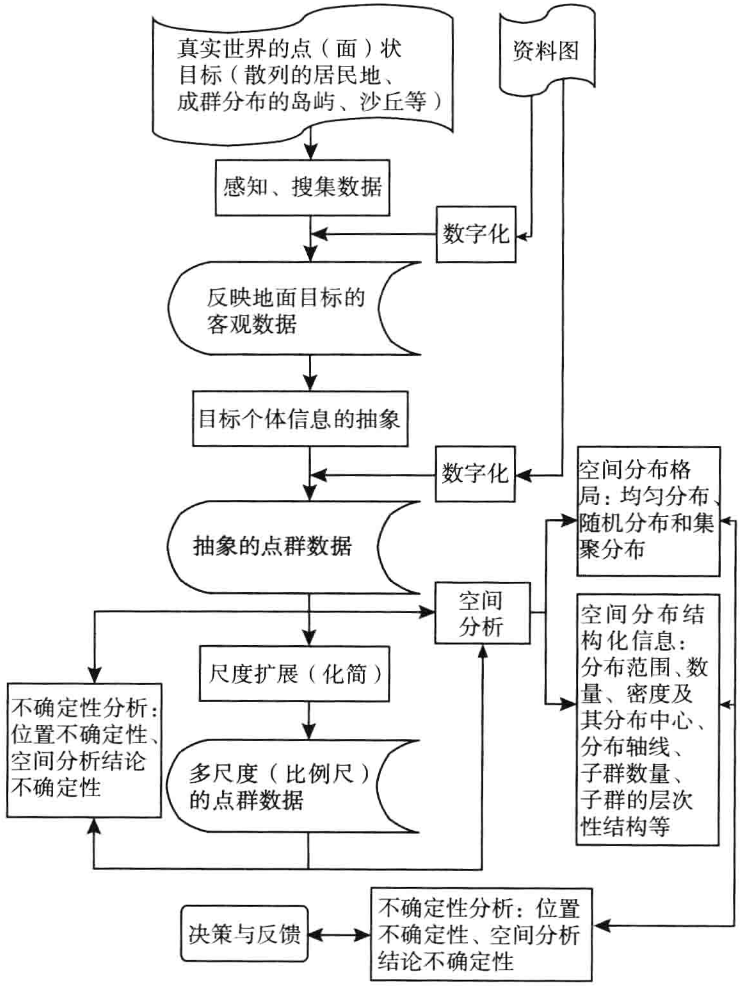
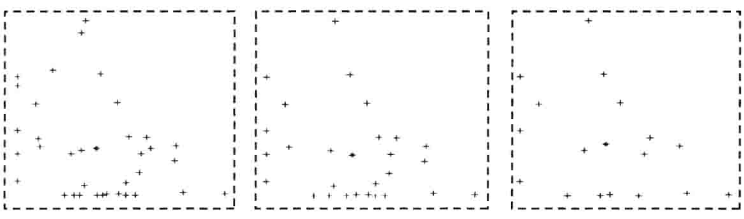


图 4.16　点群多尺度表达、空间分析、不确定性分析流程

用多个点的集合表示一个完整的地理含义，这些点可以有不同的分布规律。在地理分析中也可以把一些地理区域或多边形抽象为一些有代表性的点，然后用这些抽象的点来分析地理现象。图 4.16 体

现了点群数据的采集过程，点群数据对应于图4.16中的“抽象的点群数据”，故点群中的点所代表的要么为点状目标，要么是以点抽象的面目标。若为点状目标，其不确定性从几何意义上看，主要为位置不确定性，对点的位置不确定性已有大量的研究，如点位中误差、误差椭圆等。若是以点抽象的面目标，即使“点”成为“面”的替身，使面集的研究转变为点集的研究，即“以点带面”（图4.16中“目标个体信息的抽象”）。这个“以点带面”的过程所采用的方法虽然简单，但是初始点群数据库中抽象点群数据不确定性的主要来源应该非他莫属。因为这个“面”原来是有范围的，还有一些相关的重要属性，而在“以点带面”的过程中被忽略了，“以点带面”的过程所采用的方法不同，所得到的点群必然有所差异，进而点群所体现出来的群体特征也就不同。为了使研究的问题简化，这里不考虑由于“以点带面”给点群带来的不确定性，仅考虑图4.16中“抽象的点群数据”因为数字化而产生的位置误差及其评估方法。

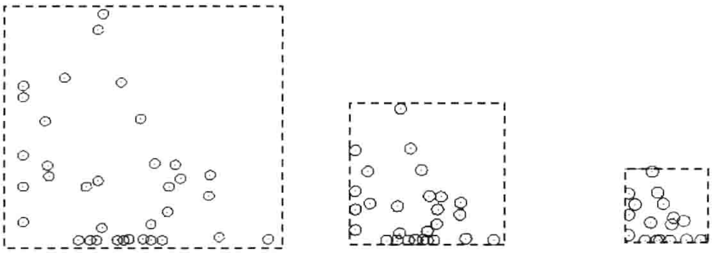
图4.17(a)是对某1：25万比例尺地形图中的点状植被进行数字化得到的数字点群，各点的标准差（$\sigma_{\text{digitizing}}$）均为 ±0.5mm 。将图4.17(a)的点群采用TIN算法分别化简为比例尺为1：50万、1：100万的数字点群如图4.17(b)，(c)所示，为了便于比较，将图4.17(b)，(c)扩展到了原始点群的大小，图4.17中“+”代表点群中的点，“＊”代表点群的算术平均中心。



（a）原始点群：1：25万　（b）尺度一：1：50万　（c）尺度二：1：100万

图4.17　点群的多尺度表达

图 4. 17 中各点的标准差用误差椭圆表示，如图 4. 18(点的坐标分量 x 与 y 值不相关，两半轴具有相同的长度)所示。采用 TIN 算法对点群的综合即为点群的选取，综合后的点群为原始点群本身的子点集，所以正如图 4. 18(b)，(c)所示，图幅面积缩小了，但是点的位置误差并没有变化。图 4. 18 反映了点群多尺度表达中各点的位置误差不随比例尺的变化而变化。



(a) 图 4.17 (a) 点位误差的误差圆　(b) 图 4.17 (b) 点位误差的误差圆　(c) 图 4.17 (c) 点位误差的误差圆

图 4. 18　图 4. 17 中点群点位误差的误差圆

图 4. 18 反映了点群中各点的位置误差，但是不能体现点群群体位置误差的大小，因此，有必要建立点群整体位置误差的指标及其量化方法。定位点的位置在其误差椭圆中的任一位置都有可能，图 4. 18 显示了点群中各点的可能位置。但是不难想象，点位中误差越大，误差椭圆就越大，所有点误差椭圆占整幅图的面积就越大，从另一个角度也说明了整幅图可能出现定位点的位置就越多，那么整个点群的位置不确定性就越大。因此，可以用所有点位误差椭圆的面积 ($S_{误差圆}$) 与整幅图面积 (S) 的比值 K 作为点群位置不确定性的评价指标，即

$$K = S_{误差圆}/S \tag{4.29}$$

对任意一组给定的点群目标 $P = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_n, y_n)\} = \{P_i(x_i, y_i), i \in N\}$ (N 取值为 $[1, n]$，n 为点群目标包含的点的个数)。假设各点的点位中误差均为 d，则

点群中每个误差椭圆的面积 $S_{误差圆i} = \pi d^2$ ，所以点位误差椭圆的面积为：

$$S_{误差圆} = \sum_{i=1}^{n} S_{误差圆i} = n\pi d^2 \tag{4.30}$$

将公式(4.30)带入公式(4.29)，得到：

$$K = n\pi d^2/S \tag{4.31}$$

点群多尺度表达实际上是对原始点群数据在不同尺度上的自动综合。点群自动综合的主要形式为：①点群的选取；②点群的典型化；③一个点群抽象为一个点；④一个点群抽象为一条线；⑤一个点群抽象为一个面；⑥点群抽象为线群；⑦点群抽象为面群。为了将问题简化，这里点群综合即为点群的选取，综合后的点群为原始点群本身的子点集。

在点群化简综合中，从大比例尺到小比例尺，开方根规律解决了选取多少点的问题，具体保留哪些点、删除哪些点由综合算法来确定，现有算法都是依据最大限度地保持空间分布格局和空间分布结构化信息来设计的。若将点群位置误差评价指标 K 引入尺度因子就能对各尺度的不确定性加以评价，并能得出点群多尺度表达位置不确定性的变化规律。

设点群获取尺度为 $1/M$ ，点位中误差为 d ，若将其化简到尺度 $1/M_A$ （ $M_A > M$ ），化简后点群中点的个数为 n_A ，图幅面积为 S_A ，则

$$K_A = S_{误差圆A}/S_A = n_A \pi d^2/S_A \tag{4.32}$$

由开方根规律，有

$$n_A = n \cdot \sqrt{M/M_A}$$

且

$$S_A = S \cdot M^2/M_A^2$$

故

$$K_A = K \cdot (M_A/M)^{3/2} \tag{4.33}$$

依据公式(4.33)，可得出：① $K_A > K$ ；② M_A 越大，K_A 就越大。因此，在点群多尺度表达中，比例尺越小，K 值越大，即随着

比例尺的减少，点群的位置不确定性增大。在具体尺度上点群位置不确定性的大小可以用指标 K 来衡量，K 的计算公式为式(4.32)和式(4.33)。

在具体的计算中，可将 K 值化为百分数或者分子为一的分数，其结果就更形象直观。利用式(4.32)计算得到图 4.18(a)的 K 值为 0.94%，利用式(4.33)得到图 4.18(b)、图 4.18(c)的 K 值分别为 2.05%、8.23%。

综上，点群整体位置不确定性可用本文建立的指标 K 来评定，K 值是一个相对量，从其计算公式来看，它适用于任意点群及其多尺度表达位置不确定性的评价。并且，从上面的分析中得到点群多尺度表达其位置不确定性的变化规律为：①点群中点的位置不确定性在点群多尺度表达中不发生变化；②点群的位置不确定性在其多尺度表达中随比例尺的缩小而增大。

3. 点群多尺度空间分析结论不确定性的评估

点群研究的核心是点群所体现的群体特征，这些群体特征由群体分布所隐含的空间结构化信息的特征参量来体现，特征参量是以点群数据计算出来的，点群数据本身的不确定性必然会传播给相应的特征参量，所以有必要对这些特征参量的准确性进行评估。这些特征参量的不确定性是一个间接计算量，满足误差传播定律。误差传播规律的详细内容可查阅相关资料。

反映点群空间分布结构化信息的特征参量有：分布范围、数量、密度及其分布、分布中心、分布轴线、子群数量、子群的层次性结构等。这些特征参量有学者给出了具体的计算方法，在这里不考虑计算这些特征参量的方法是否具有合理性，仅考虑计算结果的不确定性及其评价方法。为了将问题简化，这里所谓的点群数据均指点群中点的位置坐标。

下面以点群的算术平均中心的不确定性说明点群空间分布结构化信息不确定性的评估方法。设有 n 个离散点 P_i，其平面位置为 (x_i, y_i)，其算术平均中心为 $(\bar{x}, \bar{y})$，且 $\bar{x} = \sum x_i/n$，$\bar{y} = \sum y_i/n$，x_i，y_i 的中误差均为 d。由误差传播定律可以得出（$\bar{x}$，

$\overline{y}$）的中误差 $d_{\overline{x}}$、$d_{\overline{y}}$ 为：

$$d_{\overline{x}} = d_{\overline{y}} = d/\sqrt{n} \tag{4.34}$$

若将点群进行不同程度的化简，得到多个尺度上的点群，就其算术平均中心而言，其位置会因为计算点的个数的减少而发生变化，其中误差也会随保留点个数的减少而不断增大。若点群化简后点的个数为 n_A，则其算术平均中心的中误差 $d_{A\overline{x}}$、$d_{A\overline{y}}$ 为：

$$d_{A\overline{x}} = d_{A\overline{y}} = d/\sqrt{n_A}$$

按照式(4.34)计算得到图 4.17(a)，(b)，(c)的算术平均中心的中误差分别为±0.08mm、±0.10mm、±0.12mm，将其用误差圆表示如图 4.19(1、2、3 分别表示图(a)，(b)，(c)的算术平均中心)围绕点 1、2、3 的三个小圆。从图 4.19 可以看出，图 4.17(a)，(b)，(c)的算术平均中心并不重合，这与所选用的综合算法有关，从理论上讲，点 2、3 越接近点 1，表示所采用的算法越合理，因此，可以用不同尺度的算术平均中心到原始点群算术平均中心的距离 D 与该尺度算术平均中心的中误差（如 $d_{A\overline{x}}$、$d_{A\overline{y}}$）之和作为该尺度上算术平均中心的精度指标 d_A，即

$$d_A = D + \sqrt{d_{A\overline{x}}^2 + d_{A\overline{y}}^2}$$

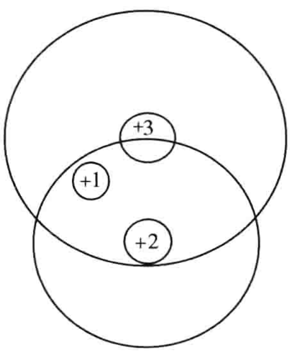


图 4.19　图 4.17 算术平均中心的误差圆

图 4.19 中分别以 2 和 3 为圆心的两个大圆即为图 4.17

(b)，(c)以式(4.29)计算得到的算术平均中心的误差圆。可见点群多尺度表达的算术平均中心的不确定性随比例尺的缩小不断增大。

点群空间分布结构化信息的其他特征参量的不确定性仍然可以按类似的方法进行评定和计算。点要素的分布模式可能随尺度的变化而改变。在小尺度下可能呈现集群分布，而在大尺度下有可能为随机分布或均匀分布，反映了点群多尺度表达随比例尺的减小其分布格局越来越模糊的特征。分析点要素任意尺度空间分布格局最常用的方法 Ripley's K 函数(Ripley，1981)，此为点位置的函数，其值也受点群位置不确定性的影响，比例尺越小，Ripley's K 函数值不确定性越大，即对点群多尺度表达分布格局的判定的不确定性越大。

从上面的分析中得到，分析任意尺度点群群体特征的空间分布格局的方法和反映点群多尺度表达空间结构化信息特征参量的不确定性随比例尺的缩小而增大。

4.5　城镇街区多尺度表达的不确定性

居民地是由人类社会生产和生活需要而形成的集聚定居地，是地图要表现的主要内容，一幅地图的总载负量中，居民地载负量占的份额最大，有时可达总量的 70% ~80%(祝国瑞等，2010)。就单个居民地而言，其内部结构、外部形状在大比例尺地图上清晰显示，在中比例尺地图上只显示概略的外围轮廓，而在小比例尺地图上只能用圈形符号表示。如图 4.20 所示，居民地的空间信息随比例尺的减小而减少，即存在性信息不断减少，不确定熵增大。因此，有必要对居民地多尺度表达的不确定性问题进行综合分析与系统研究。居民地的类型按建筑形式和分布状况可分为四类：街区式居民地、散列式居民地、窑洞式居民地以及其他类型居民地。本节主要讨论街区式居民地多尺度表达的不确定性问题。

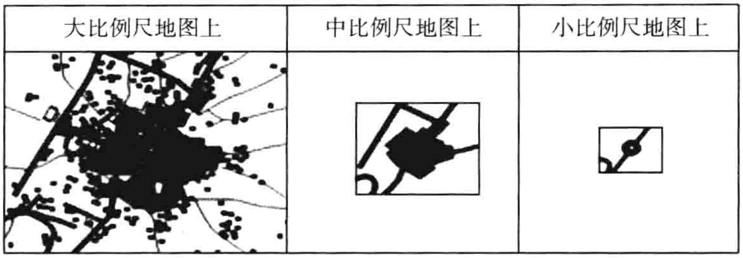


图 4.20 街区式居民地多尺度表达的图形符号(祝国瑞等，2010)

4.5.1 街区式居民地多尺度表达不确定性的本质

1. 街区式居民地多尺度表达的尺度变换过程

根据现行的测绘标准对 1 ∶ 500、1 ∶ 1000、1 ∶ 2000、1 ∶ 5000、1 ∶ 10 000、1 ∶ 25 000、1 ∶ 50 000、1 ∶ 100 000、1 ∶ 250 000、1 ∶ 500 000、1 ∶ 1 000 000 比例尺的地形图上街区式居民地表示方法的规定，街区式居民地随尺度变化有五种表现形式：①按比例尺的真实形态的缩小表达，一般由实测而得，如 1 ∶ 500、1 ∶ 1000、1 ∶ 2000、1 ∶ 5000、1 ∶ 10 000 比例尺的地形图上的居民地；②对建筑物、街区及街道进行了不同程度综合的街区式图形符号的表达，由实测地形图上居民地编绘而来；③轮廓式图形符号的表达；④圈形符号的表达；⑤街区式居民地的消失。这体现了街区式居民地在其生命期内表达的多态性。另外，依据测绘标准的规定，就单个居民地而言，在各比例尺地形图上应正确表示居民地的空间特性，包括位置、平面图形基本的形状特征、通行情况、行政意义及名称，处理好居民地与其他要素的关系。随比例尺的缩小，街区式居民地的空间特性，即位置、平面图形的基本特征、与其他要素的关系变得越来越概略。

街区式居民地在其生命期内，尺度变换的主要方法为制图综合。已有若干制图综合思想与算法可用于街区式居民地的综合，具体的综合过程发生在三个综合层面上：①发生在基本构图单元(房

屋(实地轮廓、建材质料、楼房层次、巷道、天井、庭院、临时性建筑物、街道、安全岛、绿化带、街心花园)上的连续变化，这主要发生在实测尺度上，一般不属于制图综合的范畴；②发生在由基本构图单元构成的街区、街网所表达的街区式居民地的平面图形特征(内部结构)上的连续变化，综合分两步，第一步是街区、街网的综合，第二步是街区内部地物的综合，包括街道与建筑物的移位处理和建筑物的综合；③发生在由街道网、街区和其他各种建筑物的分布范围构成的街区式居民地外部形状特征上的连续变化，重点综合其轮廓特征以及与其他要素的关系。在每个综合层面上发生的是数据量上的变化(也可以将这三个综合层面理解为三个尺度变化区间)，在从第一个尺度变化区间过渡到第二个尺度变化区间、第二个尺度变化区间过渡到第三个尺度变化区间、第三个尺度变化区间过渡到用圈形符号表示、用圈形符号表示到消亡(不表示)的过程中，居民地发生的是数据质上的变化(突变)，也可以将这些过渡点称为尺度变化点。街区式居民地尺度变换区间、尺度变换点及各区间内变换的内容如图4.21所示。

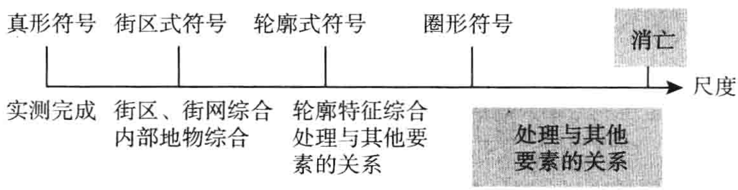


图4.21　生命期内街区式居民地尺度变换过程

2. 街区式居民地多尺度表达不确定性的本质

随街区式居民地表现形式的变化，其综合变换的目标变得越来越宏观，这就从根本上决定了在街区式居民地多尺度表达的过程中，其空间特性变得越来越概略，其不确定性不断增大。图4.21挖掘了存在这一变化的根本原因是由于制图综合而产生的。

胡圣武(2006)认为不确定性是关于空间过程和特性不能被准确确定的程度，是自然界各种空间现象本身固有的属性。因此，街

区式居民地多尺度表达的不确定性是其空间特性不能被准确确定的程度，也就是其位置、平面图形基本形状特征、通行情况、与其他要素的关系随比例尺的变化不能被准确确定的程度，即位置不确定性、平面图形基本形状特征的不确定性、通行情况的不确定性、与其他要素关系的不确定性。这就是街区式居民地多尺度表达不确定性的本质。这里的位置不确定性包括居民地内部各点位、地物的位置不确定性，外围轮廓的位置不确定性，建筑物移位的位置不确定性；平面图形基本形状特征的不确定性包括内部结构特征的不确定性和外围轮廓特征的不确定性；通行情况的不确定性主要讨论街区内部通行情况的不确定性；与其他要素的关系的不确定性主要指居民地与河流、铁路、高速公路等的联系的不确定性，也可将其理解为居民地与外部地物的拓扑关系的不确定性。关于街区式居民地多尺度表达包含的不确定性的类型可用图 4.22 表示。

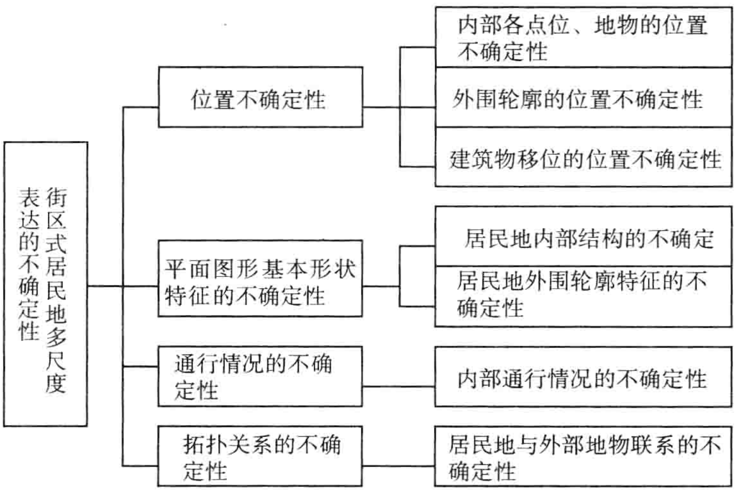


图 4.22　街区式居民地多尺度表达不确定性的类型

综合分析图 4.21 和图 4.22，可以概括出街区式居民地多尺度表达各尺度阶段的不确定性类型，如图 4.23 所示。

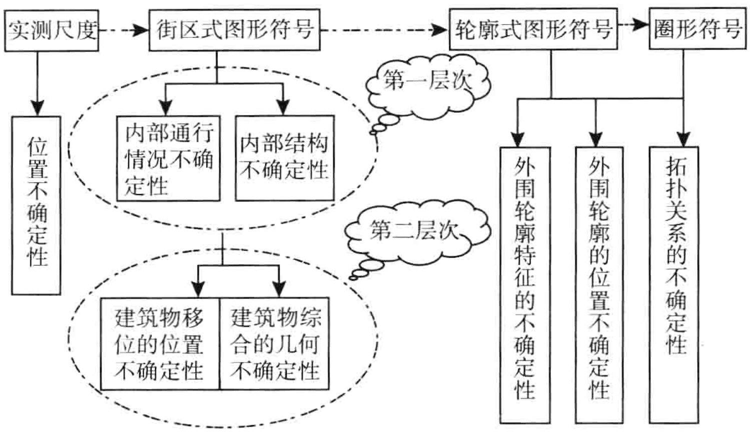


图4.23　街区式居民地多尺度表达各尺度阶段不确定性类型

4.5.2　街区式居民地多尺度表达不确定性评价方法

在实测尺度阶段，其不确定性一般考虑居民地内各重要点位及地物的位置不确定性(测量误差)。在用轮廓式图形符号表达阶段，外围轮廓的综合主要是轮廓线的化简及处理与其他要素的联系问题。轮廓线化简引起的不确定性有外围轮廓位置的不确定性、外围轮廓特征的不确定性，处理与其他要素联系时引起的不确定性为拓扑关系的不确定性，这三类不确定性可以利用已有的相关研究成果进行评价。

街区、街网综合阶段，街区式居民地用街区式图形符号表示，平面图形特征发生连续变化，综合为两步：第一步街区、街网的综合，其不确定性对应于图4.23中的第一层次：内部通行情况和内部结构的不确定性；第二步街区内部地物的综合，其不确定性对应于图4.23中的第二层次：建筑物位移的位置不确定性和建筑物综合的几何不确定性。

1. 街区、街网综合的不确定性评价方法

在街区、街网综合阶段，主要是对居民地内部结构特征的综

合，街道网图形是居民地内部结构的骨架元素，主要通过街网的综合(体现在短小街道的舍弃)实现街区的合并，或通过小街区的合并达到短小街道的舍弃来实现街区、街网的综合(毋河海，2000)，然后通过删除的方法综合街区的凹凸拐角。街网图形的简化引发街区式居民地内部结构及通行情况的变化，引发的不确定性也就是居民地内部结构的不确定性及其内部通行情况的不确定性。

街区、街网综合是在整体构思、科学原则的指导下进行的，综合的宗旨即是在保证居民地内部通行情况的条件下保持街道网的相似性。街道网的组成单元为街巷，对于街道网图形结构，可以从街巷的总长度、街道网的密度、各分区街道网密度对比、街道网的形状特征等方面对街道网图形结构进行描述。从街区的角度来看，可用街区个数、街区单元最大面积、街区单元最小面积等来描述居民地。因此，综合前后这些指标的变化就是街区式居民地内部结构不确定性评价的具体内容，具体指标见表4.2。

表4.2中街道总长度（L_A 、L_B）、街区数量（M_A 、M_B）、街区单元最大面积($A_{\max A}$ 、$A_{\max B}$)、街区单元最小面积($A_{\min A}$ 、$A_{\min B}$)可以从图上或者数据库中直接统计或量算出来，街道网密度(η_A 、η_B)可用街道总长度（L_A 、L_B）与居民地面积 S 的比值进行量化，即 $\eta_A = L_A/S$ 、$\eta_B = L_B/S$ 。街网综合的宗旨即是在保证居民地内部通行情况的条件下保持街道网的相似性，因此，综合前后街道网的形状特征变化可用街道网的几何相似度来衡量，但是关于多尺度空间数据相似关系模型与计算方法的研究在国内还处于起步阶段(闫浩文等，2006)，如何计算街道网的几何相似度也有待于进一步研究。对于街道网的密度分区，在综合前后均可以采用聚类的方法将街道网分成若干区域，计算各区域街道网密度，得到密度对比关系；然后将综合前后的分区及密度对比关系进行比较，得到综合前后街道网密度分区及对比关系的变化特征。

关于通行情况的不确定性问题，可以借用道路通行能力计算的相关方法加以计算并进行评价。

表4.2　　**街区式居民地内部结构不确定性评价方法**

<table>
<tr><th>不确定性评价内容</th><th>评价指标</th><th>计算方法</th><th>公式</th><th>不确定性大小</th></tr>
<tr><td>街巷总长度的变化</td><td>街道总长度的减少量 ΔL</td><td>新图上街道长度 L_B 与原图上街道长度 L_A 的差值</td><td>$\Delta L = L_B - L_A$</td><td rowspan="5">指标值越小，综合程度越大，不确定性越大</td></tr>
<tr><td>街道网的密度变化</td><td>街道网密度的变化 K_η</td><td>新图上街道网密度 η_B 与原图街道网密度 η_A 的比值</td><td>$K_\eta = \frac{\eta_B}{\eta_A}$</td></tr>
<tr><td>街区数量变化</td><td>街区数量比 K_M</td><td>新图上街区个数 M_B 与原图上街区个数 M_A 的比值</td><td>$K_M = \frac{M_B}{M_A}$</td></tr>
<tr><td>街区单元最大面积变化</td><td>街区单元最大面积差值 ΔA_{max}</td><td>新图上街区单元最大面积 ΔA_{maxB} 与原图上街区单元最大面积 ΔA_{maxA} 的差值</td><td>$\Delta A_{max} = \Delta A_{maxB} - \Delta A_{maxA}$</td></tr>
<tr><td>街区单元最小面积变化</td><td>街区单元最小面积差值 ΔA_{min}</td><td>新图上街区单元最小面积 ΔA_{minB} 与原图上街区单元最大面积 ΔA_{minA} 的差值</td><td>$\Delta A_{min} = \Delta A_{minB} - \Delta A_{minA}$</td></tr>
<tr><td>街道网的形状特征变化</td><td>相似度</td><td>新图上街道网的形状特征与原图的相似程度</td><td>有待于进一步研究</td><td></td></tr>
<tr><td colspan="3">街道网密度分区及对比关系的变化</td><td colspan="2">用聚类方法进行分区，用上文提到的方法计算各分区密度</td></tr>
</table>

2. 建筑物综合不确定性的评价方法

在建筑物综合阶段，采用的综合手段为建筑物合并、建筑物删除、建筑物轮廓图形概括。在该阶段综合时重点围绕建筑物进行，综合过程基本上不影响居民地的位置、通行情况以及与其他要素的

关系，只是引起平面图形基本形状特征细微的变化和建筑物的形状、建筑面积等的变化，与之对应的不确定性即为居民地内部结构的不确定性。另外，因为道路符号要宽于真实比例尺下显示的地面道路，从而产生道路间、道路和建筑物的叠置，综合时通过仿射变换(移位、简化、合并等)的方法来解决，现有研究中有不少关于街道与建筑物冲突的矢量移位模型，这带来了建筑物和街道因为移位而产生的位置不确定性，可以采用目标位移的数量、位移的最大值、中位数、标准差等来评定。

建筑物合并的结果是建筑物数量减少且合并后的图形面积大于合并前各分离图形面积之和；建筑物删除的结果是建筑物数量减少且删除后图形面积小于原图形面积；建筑物轮廓图形概括主要是删除其凹凸轮廓以及天井和庭院，结果是建筑物变得规则，图形面积可能增加，也可能减少。因此，该阶段产生的不确定性一是建筑物数量变化，建筑物个数明显减少；二是建筑面积的变化，是建筑物合并、删除、轮廓概括共同导致的结果；三是建筑物轮廓特征的变化，建筑物轮廓变得更加规则。从综合结果的整体上来看，除了建筑物数量、建筑面积、建筑物轮廓特征的变化之外，就是建筑物密度的变化，由建筑面积变化导致，另外居民地内部各区域建筑物密度对比关系也可能发生变化。这时综合前后建筑物的数量、建筑面积、建筑密度、建筑物的轮廓特征、建筑密度分区的变化就是不确定性评价的内容。建筑物的数量、建筑面积、建筑密度在地图上或是地图数据库中都是非常容易量算的，与它们对应的不确定性评价指标由表4.3进行详细定义。

表4.3中建筑物个数(N_A 、N_B)、建筑物面积(S_A 、S_B)可以从图上或者数据库中直接统计或量算出来，建筑密度(λ_A 、λ_B)可用建筑面积(S_A 、S_B)与居民地面积 S 的比值进行量化，即 $\lambda_A = S_A/S$ 、$\lambda_B = S_B/S$ 。建筑物轮廓特征与建筑密度分区在图上或地图数据库中难以直接评估和量化。对于建筑物轮廓特征，综合程度越高，其复杂程度越低，轮廓特征越规则。在数字图像领域，关于图形复杂程度的描述，常用的数量指标是分维数，分维数越大，图形越复杂。这里引用图形分维数来描述建筑物轮廓特征的规则性，用

表 4.3　**街区式居民地建筑物综合不确定性的评价方法**

<table>
<tr><th>不确定性评价内容</th><th>评价指标</th><th>计算方法</th><th>公式</th><th>不确定性大小</th></tr>
<tr><td>建筑物数量的变化</td><td>建筑物数量比 K_N</td><td>新图上建筑物个数 N_B 与原图上建筑物个数 N_A 的比值</td><td>$K_N = \frac{N_B}{N_A}$</td><td rowspan="4">指标值越小，综合程度越大，不确定性越大</td></tr>
<tr><td>建筑面积的变化</td><td>建筑面积减少量 ΔS</td><td>新图上建筑物面积 S_B 与原图上建筑物面积 S_A 的差值</td><td>$\Delta S = S_B - S_A$</td></tr>
<tr><td>建筑密度的变化</td><td>总建筑密度变化 K_λ</td><td>新图上建筑密度 λ_B 与原图建筑密度 λ_A 的比值</td><td>$K_\lambda = \frac{\lambda_B}{\lambda_A}$</td></tr>
<tr><td>建筑物轮廓特征的变化</td><td>建筑物规则性差异 ΔD</td><td>新图上建筑物图形分维数 D_B 与原图上建筑物图形分维数 D_A 的差值</td><td>$\Delta D = D_B - D_A$</td></tr>
<tr><td colspan="2">建筑物密度分区及对比关系的变化</td><td colspan="3">用聚类方法进行分区，用上文提到的方法计算各分区密度</td></tr>
</table>

建筑物规则性差异作为建筑物轮廓特征变化不确定性的评价指标，详细的定义见表 4.3，关于图形分维数的计算可以查阅相关资料。对于建筑物的密度分区，在综合前后均可以采用聚类的方法将居民地分成若干区域，计算各区域建筑密度，得到各分区密度对比关系；然后将综合前后的分区及密度对比关系进行比较，得到综合前后建筑物密度分区及对比关系的变化特征。

4.6　本章小结

本章重点研究了各类空间数据多尺度表达几何不确定性的评价指标和量化模型。第一，发展了点元多尺度表达位置不确定性的评

价指标和量化模型。第二，在充分分析和研究线状要素多尺度表达方法和线状要素多尺度表达不确定性基本问题的基础上，研究了线状要素多尺度表达位置不确定性的评价方法，构建了评价指标和量化模型，研究了线状要素多尺度表达形状不确定性的评价方法。第三，研究了面状要素多尺度表达不确定性的评价方法，由于面的边界由线来界定，因此，为了避免重复，这里只研究了多边形多尺度表达面积精度的评价方法。第四，研究了点群多尺度表达位置不确定性和多尺度点群空间分析结论的不确定性问题，并建立了相关评价方法。第五，研究了街区式居民地多尺度表达不确定性的评价方法，建立了相关评价指标和量化模型。

第5章　空间数据多尺度表达的空间关系不确定性建模

GIS空间关系主要分为顺序关系、度量关系、拓扑关系三大类。拓扑空间关系是指拓扑变换下的拓扑不变量，如空间目标的相邻和连通关系，以及表示线段流向的关系等。顺序关系是指描述目标在空间中的某种排序，如前后、上下、左右、东西南北等。度量关系是用某种度量空间中的度量来描述的目标间的关系，如目标间的距离、目标相分割的程度等。

由于地理目标自身的不确定性，空间数据的获取、处理的不确定性以及人类对空间关系的认知的不确定性，必然导致地理空间目标之间空间关系的不确定性。在空间数据库中，空间关系是利用某空间关系表达模型通过空间数据直接计算出来的，空间数据存在不确定性，空间关系也就必然存在不确定性。如果尺度发生变化，空间数据本身的不确定性将得到传播，并与数据尺度转换过程中引入的不确定性一起构成尺度转换后数据的不确定性，由此计算出的空间关系可能也会不同。

本章以空间数据多尺度表达的不确定性为基础，利用粗集理论，分析空间数据多尺度表达空间拓扑和方向关系的不确定性，并构建空间数据多尺度表达空间拓扑和方向关系不确定量的评价模型。

5.1　多尺度空间面目标的粗集表达

这里结合2.2.2节的内容（不确定性地理目标模型），定义空间面目标尺度扩展后的粗集表达。

对于给定点位误差的平面 $A(x_1,y_1,x_2,y_2,\cdots,x_m,y_m,x_1,y_1)$，假

设面 A 误差环内边界线方程为 $F_{A0}(y)$ 、外边界线方程为 $F_{A1}(y)$ ；由 $F_{A0}(y)$ 所包围的平面区域为 $S(F_{A0}(y))$ ，由 $F_{A1}(y)$ 所包围的平面区域为 $S(F_{A1}(y))$ ，则平面 A 的下近似为 $\underline{A}=S(F_{A0}(y))$ 、上近似为 $\overline{A}=S(F_{A1}(y))$ 。如图 5.1(a)所示。

尺度扩展时，面目标按要求进行化简，简化后面目标的不确定性值可按 4.3 节(面状要素多尺度表达的不确定性)计算出来，并可建立其误差环及误差环的边界线方程。因此，按照不确定目标粗集表达的方法，可以得出任意尺度上空间目标位置不确定性的粗集模型。图 5.1(b)是图 5.1(a)简化的结果，设 5.1(b)的内边界线方程为 $F_{A'0}(y)$ 、外边界线方程为 $F_{A'1}(y)$ ，则图 5.1(b)的平面 A' 的下近似为 $\underline{A'}=S(F_{A'0}(y))$ 、上近似为 $\overline{A'}=S(F_{A'1}(y))$ 、边界为 $\mathrm{bn}(A')=\overline{A'}-\underline{A'}$ 。

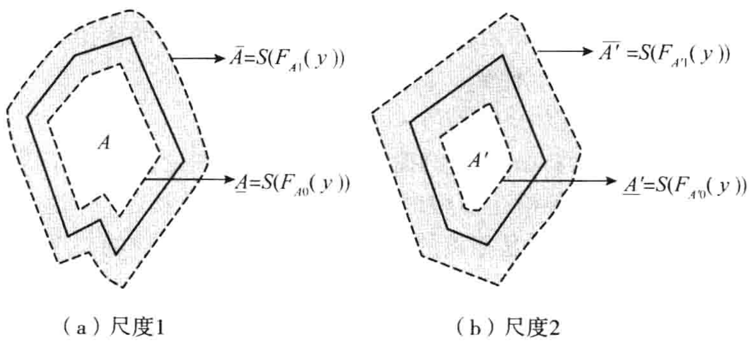


图 5.1　多尺度面目标位置不确定性的粗集表达

5.2　多尺度空间拓扑关系不确定性建模

作为一种基本空间关系，拓扑关系是空间数据表达和定性空间推理研究的基本问题之一。目前，对空间拓扑关系的描述主要通过两种途径进行，一是基于集合理论的数学形式拓扑关系模型(4-交模型和 9-交模型)，二是基于逻辑推理的公理化方法(Cohn & Hazarin,

2001)，如 RCC-5 和 RCC-8。

拓扑关系的描述无论采用上述哪种途径，都是以地理目标对应的空间数据为基础进行运算或推理的。由于空间数据具有不确定性，拓扑关系自然也存在不确定性。目前，已有若干描述不确定区域之间拓扑关系的模型，如 Cohn 等人(1996)建立的“卵-黄”模型等。

如今，空间数据的多尺度表达备受关注。空间数据在多尺度表达的过程中，其相应的拓扑关系也得到了适当程度的综合，拓扑关系随之发生了变化。从这个意义上来说，空间数据多尺度表达中，拓扑关系的不确定性增大了。因此，我们以现有的拓扑关系模型和不确定拓扑关系模型为基础，发展了多尺度空间拓扑关系不确定性评价模型。

5.2.1　空间拓扑关系表达模型

本小节对基本拓扑关系及其度量扩展模型进行阐述和分析，并在拓扑关系概念邻域和拓扑距离的基础上，定义扩展的拓扑距离。

1. 基本拓扑关系模型

(1)RCC 及 RCC-D 模型

RCC 模型是 Clarke(1981)提出的拓扑关系理论，当两个区域拥有相同的部分，或者相互接触时，这两个区域是连接的，即为 $C(x,y)$；否则，这两个区域是相离的，即为 $\neg C(x,y)$。相应的公理如下：

公理 1：$\forall x C(x,x)$。对于任意的(表达式中“$\forall$”可以读成“任意的”，下同)区域 x，其自身总是连接的，即区域连接关系具有自反性。

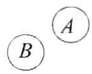
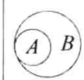
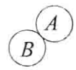
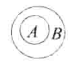
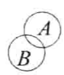
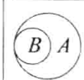
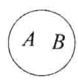
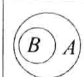
公理 2：$\forall x \forall y(C(x,y) \Rightarrow C(y,x))$。对于任意的区域 x 和 y，如果 x 与 y 连接，则 y 与 x 也必然连接(表达式中的“$\Rightarrow$”符号可以读成“推出”，下同)，即区域连接关系具有对称性。

公理 3：$\forall x \forall y(\forall z C(z,x) \Leftrightarrow C(z,x) \Rightarrow x = y)$。对于任意的区域 x、y、z，如果当且仅当 z 与 y 连接时，z 与 x 连接(表达式中的“$\Leftrightarrow$”符号可以读成“蕴涵”，下同)，则 x 与 y 相等。

上述三个公理构成了基于逻辑的空间拓扑关系描述及推理模型的基石。基于上述三个关系的逻辑定义衍生出的目标之间的拓扑关

系集合为{DC,EC,PO,EQ,TPP,NTPP,TPP^{-1},$NTPP^{-1}$}（“-1”表示相对称的一种拓扑关系），这个集合在理论上被称为 RCC-8。它描述了任意两个区域之间可能的拓扑关系，即 DC（相离）、EC（邻接）、PO（重叠）、EQ（相等）、TPP（内切于）、NTPP（包含于）、TPP^{-1}（被内切），$NTPP^{-1}$（被包含），见表 5.1。这 8 个关系间不相交且联合完备，即组成 JEPD 集（jointly exhaustive and pair wise disjoint，互不相交且联合完备集）。

表 5.1　　**RCC-8 的 8 种基本关系**

谓词	语义	图示	谓词	语义	图示
DC(x,y)	相离		TPP(x,y)	内切于	
EC(x,y)	邻接		NTPP(x,y)	包含于	
PO(x,y)	重叠		TPP^{-1}(x,y)	被内切	
EQ(x,y)	相等		$NTPP^{-1}$(x,y)	包含	

如果不考虑区域的边界，RCC-8 变为 RCC-5，如图 5.2 所示，即不能区分 DC 和 EC、TPP 和 NTPP、TPP^{-1} 和 $NTPP^{-1}$。这些关系分别合并后，DC 和 EC 变成 RCC-5 中的不连续关系 DR，TPP 和 NTPP 变成 RCC-5 中的包含关系 PP，TPP^{-1} 和 $NTPP^{-1}$ 变成 RCC-5 中的包含于 PP^{-1}。这样，RCC-5 包含的 5 种关系为：DR、PO、PP、EQ 和 PP^{-1}。这 5 个关系组成 JEPD 集。

RCC 模型本身是定义在连续区域上的，许多学者对其进行了扩展，Galton（1999）基于隶属和连接关系定义了基于图的离散空间中的 RCC，即 RCC-D。离散空间中 RCC-D-5 对应于 RCC-5，RCC-D-8 对应于 RCC-8。RCC-D 是研究基于图的离散空间中的拓扑关系的一个重要参考。由于 RCC-5 和 RCC-D-5 不能区分两个地理对象之

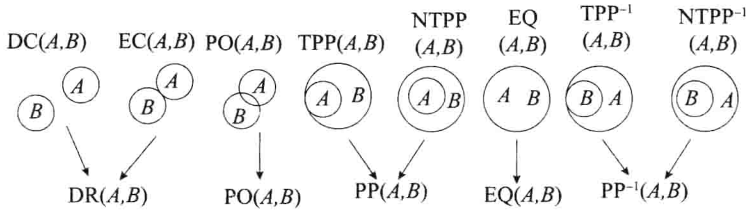


图5.2　RCC-8和RCC-5的关系及其对应

间的相接与相离、真部分与正切真部分的关系，而实际的一些应用中可能会需要对这些关系进行描述，例如，山谷和山脊的关系、地籍块的相邻关系等。一般认为RCC-8和RCC-D-8为研究空间拓扑关系的标准JEPD集。

(2)4-交模型

4-交模型是Egenhofer等人(1991)提出的一种基于点集理论的空间拓扑关系分类模型，借助于点集拓扑理论，空间目标A可分为两部分：内部和边界，分别记作A^0和∂A。目标之间的拓扑关系能够用内部和边界的4-交模型来描述，记作：

$$R_4 = \begin{pmatrix} \partial A \cap \partial B & \partial A \cap B^0 \\ A^0 \cap \partial B & A^0 \cap B^0 \end{pmatrix}$$

通过这些几何的交集是否为空来判断空间目标A和B之间的拓扑关系，理论上，共有$2^4=16$种拓扑关系。在实际应用中，有些空间拓扑关系是不存在的，面与面之间有8种拓扑关系存在，线与线之间有16种空间拓扑关系，线与面之间有11种空间拓扑关系。

(3)9-交模型

9-交模型是在4-交模型的基础上发展起来的，用于描述两个地理空间目标之间的空间拓扑关系的三个集合是：目标的内部、目标的边界、目标的外部。目标的内部、边界的定义都与4-交模型中的定义相同，目标的外部可被定义为目标集合的补集，因为该目标是位于一个有限的空间中。设有两个空间目标A和$B(A \neq \varnothing, B \neq \varnothing)$，它们的内部、边界、外部分别为：$A^0$、$\partial A$、$A^-$；$B^0$、$\partial B$、$B^-$，通过这6个部分的交集，9-交模型为：

$$R_9 = \begin{pmatrix} A^0 \cap B^0 & A^0 \cap \partial B & A^0 \cap B^- \\ \partial A \cap B^0 & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^0 & A^- \cap \partial B & A^- \cap B^- \end{pmatrix}$$

9-交模型能够区分的拓扑关系类型为2种点/点关系、3种点/线关系、3种点/面关系、33种线/线关系、19种线/面关系及8种面/面关系。

(4)4-交差模型

4-交差模型(4-intersection-difference model,简称4ID)是邓敏等人(2005)提出的,它是由两个面目标 A 的内部与 B 的内部之交集 $(A^0 \cap B^0)$、A 的边界与 B 的边界之交集 $(\partial A \cap \partial B)$、$A$ 与 B 之差集 $(A - B)$ 以及 B 与 A 之差集 $(B - A)$ 四个部分构成的,形式化表达为:

$$R_{4ID}(A,B) = \begin{bmatrix} A^0 \cap B^0 & A - B \\ B - A & \partial A \cap \partial B \end{bmatrix} \tag{5.1}$$

或等价表达为一个四元组形式,即

$$R_{4ID}(A,B) = [A^0 \cap B^0, A - B, B - A, \partial A \cap \partial B] \tag{5.2}$$

如果式(5.1)或(5.2)中各元素取值为空($\varnothing$)与非空($\neg\varnothing$)时,则有 $2^4 = 16$ 种可能取值,即4-交差模型能够区分16种拓扑情形。由于简单面目标几何构成的特殊性(例如,连续的内部和外部),使得在一定分类层次上仅有8种拓扑关系具有物理意义。

上面分别介绍了4种最基本的空间拓扑关系表达模型,RCC-8、9-交模型、4-交模型、4-交差模型均能区分8种面/面拓扑关系,计算模型虽然不一样,但它们能够描述和区分的8种拓扑关系是一致的。在计算简单面目标间的拓扑关系时,用这4种模型中的任意一种都是可以的。

2. 基本拓扑关系模型的扩展

1)边界交集成分描述模型

为了详细描述复杂形状区域之间的拓扑关系,Egenhofer 等人(1994)提出了一种用边界交集成分来描述区域之间拓扑关系的方法。边界交集成分描述模型是在4-交集模型描述的基础上,对两

个目标之间边界相交情况的一种细化。若两个目标的边界有不同的相交成分，则将这些相交的部分称为边界交集成分。边界交集成分用成分系列(component sequence)及其维数(component dimension)、类型(component type)和有界性(boundedness)等性质来描述。

①成分序列。沿一个区域的边界，按照顺时针(或者逆时针)方向依次记录这些成分，形成一个递增的数字序列。

②成分维数。两个区域的边界和边界相接形成的交集成分称为相接成分，如果这个交集为一个点，则称为0维的相接成分；如果交集为一条线，则称为1维的相接成分。

③成分类型。两个区域的边界和边界相交形成的交集成分称为交叉成分。若相交成分从同一个部分进入和离开，称为相接触(touch，简写为t)，相接触成分可能发生在重叠、邻接、被内切和内切于关系的两区域之间；若相交成分进入和离开的位置不同，称为交叉(cross，简写为c)，交叉成分仅仅适用于相交(overlap)的区域，如果一个区域的边界从另一个区域的外部进入其内部，相应的成分就具有“交叉进”(crossing into，简写为ci)的特性；反之则为“交叉出”(crossing out，简写为co)。如果一个1维交叉成分在两个区域并集的内部，则该成分具有“内部交叉”(inner crossing，简写为ic)的特性；如果在并集的边界上，则具有“外部交叉”(outer crossing，简写为oc)的特性。

④有界性。那些与无界外部(两个区域并集的补集的无边界的子集)邻近的边界成分称为无界成分(unbounded components，简写为u)；否则称为有界成分(bounded components，简写为b)。

两个区域详细的拓扑关系用一个4-交矩阵$M(A, B)$来表达，它描述内容不变量，在非空边界-边界交集的情况下，列出边界-边界成分序列和每个成分维数、类型和有界性的成分不变量表$T(\partial A, \partial B)$。例如，图5.3的相交结构被描述为4-交矩阵(式5.3)和组件不变量表(式5.4)。

$$M(A,B)=\begin{pmatrix}\neg\varnothing & \neg\varnothing \\ \neg\varnothing & \neg\varnothing\end{pmatrix} \tag{5.3}$$

$$T(\partial A,\partial B)=\left[\begin{array}{c|ccccccc} \text{sequence} & 1 & 5 & 6 & 7 & 4 & 3 & 2 \\ \hline \text{dimension} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{type} & \text{ico} & \text{cl} & \text{co} & \text{cl} & \text{co} & \text{t} & \text{cl} \\ \text{boundedness} & \text{u} & \text{u} & \text{u} & \text{u} & \text{b} & \text{b} & \text{b} \end{array}\right] \tag{5.4}$$

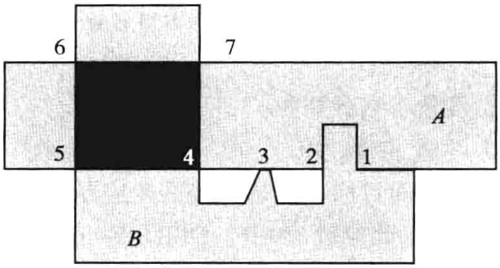


图 5.3　一个详细的拓扑关系

2）面目标（区域）间基本拓扑关系的度量参数

RCC-8、4-交模型、9-交模型等都能区分面目标间的 8 种基本拓扑关系，但是，这 8 种基本拓扑关系过于粗略，邓敏等人（2002、2009）、郭庆胜等人（2005）对区域与区域之间的空间关系进行了扩展，建立了区域与区域之间空间关系的量化描述（度量分析）方法，以满足空间拓扑关系精化描述的需要。

（1）简单区域与区域间基本拓扑关系的度量参数

邓敏等人（2002）基于 9-交模型以 A 为参照目标，B 为空间查询和分析对象，构建了点集分割度量参数：IIb（表达 B 的内部被 A 的内部分割的程度）、BBb（表达 B 的边界被 A 边界分割的程度）和点集接近性度量参数：内接近性度量（ICb）、外接近性度量（ECb）来辅助描述区域与区域之间的拓扑关系。这 4 个参数的取值范围都不是在区间[0,1]上。邓敏等人（2009）为了将度量空间关系转换为自然语言空间关系，将上述 4 个度量参数进行了改进，基于 4-交差模型定义了 4 个分割度量参数、3 个接近性度量参数来精化描述基本拓扑关系。4 个分割度量参数分别为：

公共边界长度分割比：CB = Length ($\partial A \cap \partial B$) /Length (X) ，其中，$X \in (\partial A, \partial B, \partial A \cup \partial B)$ ；

面积差值比：DZ = Area ($A - B$) /Area (X) ，其中，$X \in (A, B, A \cup B)$ ；

面积反差值比：SDZ = Area ($B - A$) /Area (X) ，其中，$X \in (A, B, A \cup B)$ ；

公共区域面积比：CA = Area ($A^\circ \cap B^\circ$) /Area (X) ，其中，$X \in (A, B, A \cup B)$ 。

3 个接近性度量参数分别为：

内接近性度量参数 (IC)：IC = 1 − Area ($B \Theta\ d_{\min}(B, A)$) / Area (B)，用来表达面目标 B 的边界接近 A 边界的程度；

分离式接近性度量参数 (DEB)：DEB = 1 − Area (B) /Area ($B \cdot d_{\min}(\partial B, \partial A)$)，用来表达面目标 B 的边界从 A 的外部接近 A 边界的程度；

包含式外接近性度量参数 (CEC)：CEC = 1 − Area (B) /Area ($B \cdot d_{\min}(\partial B, \partial A)$)，用来表达面目标 B 的边界从 A 内部接近 A 边界的程度。

上面 7 个度量参数的表达式中，· 、Θ 分别为形态膨胀和收缩算子；$d_{\min}(\partial B, \partial A)$ 为目标 B 的边界到目标 A 的边界的最短距离，计算表达式为 $d_{\min}(\partial B, \partial A) = \min(d_{\min}(p, \partial A) \mid p \in \partial B) = \min(d(p,q) \mid p \in \partial B, q \in \partial A)$ 。

CB、DZ、SDZ、CA、IC、DEB、CEC 这 7 个度量参数的取值范围皆为(0,1)。每种拓扑关系与度量分割参数和接近性度量参数的对应关系见表 5.2。

表 5.2　**拓扑关系类型与度量参数的对应关系 (邓敏等，2009)**

基本拓扑关系	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 $NTPP^{-1}$	内切于 TPP	包含于 NTPP	相等 EQ
度量参数	DEB	CB	CB、CA	CB、CA	CA、CEC	CB、CA	CA、IC	

分析表5.2发现,DZ、SDZ并未与任何拓扑关系对应,而且4-交差模型计算得到的简单区域之间拓扑关系的类型与RCC-8、4-交模型、9-交模型的结果是一致的,因此,CB、CA、IC、DEB、CEC可以作为8种区域与区域之间拓扑关系的度量参数。IC是专门针对"包含于"关系建立的、DEB是专门针对"相离"关系建立的、CEC是专门针对"包含"关系建立的。对CA进一步分析发现,CA在相离或邻接关系时始终等于0,在被内切、相等、包含关系时始终等于1,在内切于、包含于关系时是个固定值。因此,CA仅对重叠关系的比较有意义。对CB做进一步分析,它对于度量邻接、重叠、被内切、内切于关系都有意义,但是,此处对于重叠关系,假定区域重叠,而边界仅有公共的点,即重叠关系的度量参数不考虑CB。因此,表5.2可简化为表5.3。

表5.3　**拓扑关系类型与度量参数的对应关系**

基本拓扑关系	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 $NTPP^{-1}$	内切于 TPP	包含于 NTPP	相等 EQ
度量参数	DEB	CB	CA	CB	CEC	CB	IC	

以*A*为参照目标,*B*为源目标,与基本拓扑关系对应的度量参数的变化特征见表5.4。

表5.4　**度量参数特征**

基本拓扑关系	度量参数	取值范围	变化趋势(以*A*为参照目标,*B*为源目标进行分析)
相离 DC	DEB	[0,1]	*A*距离*B*越远,DEB越大;*A*距离*B*无穷远时,DEB=1;*A*距离*B*为0时,DEB=0
邻接 EC	CB	[0,1]	公共边界为一点时,CB=0;公共边界越长,CB越大;*B*的边界完全为公共边时,CB=1

续表

基本拓扑关系	度量参数	取值范围	变化趋势(以 A 为参照目标,B 为源目标进行分析)
重叠 PO	CA	[0,1]	公共区域为一点时,CA=0; B 与 A 相交面积越大,CA 越大; B 与 A 完全相交时,CA=0
被内切 TPP^{-1}	CB	[0,1]	公共边界为一点时,CB=0;公共边界越长,CB 越大; B 的边界完全为公共边时,CB=1
包含 $NTPP^{-1}$	CEC	[0,1]	A 的边界距离 B 的边界为 0 时,CEC=0; A 的边界距离 B 的边界越远,CEC 越大; B 位于 A 的中心,且为一点时,CEC=1
内切于 TPP	CB	[0,1]	公共边界为一点时,CB=0;公共边界越长,CB 越大; B 的边界完全为公共边时,CB=1
包含于 NTPP	IC	[0,1]	A 内切于 B 时,IC=1; A 包含于 B 时,A 的边界距离 B 的边界越远 ,IC 越小; A 位于 B 的中心,且为一点时,IC=0

(2)复杂区域边界交集成分度量关系

郭庆胜等人(2005)为了适应拓扑关系抽象的需要,引入度量关系以精确描述两个区域的边界交集成分。具体用成分关联区域度量特征表(表 5.5)和 1 维成分的长度(表 5.6)来描述两个区域的边界交集成分。

表 5.5　**成分关联区域度量特征表**

成分关联区域	(1,2)	(2,3)	(3,4)	…	(n,l)
面积	A_{12}	A_{23}	A_{34}	…	A_{nl}
长边界长度	le_{12}	le_{23}	le_{34}	…	le_{nl}
短边界长度	se_{12}	se_{23}	se_{34}	…	se_{nl}
扁率	f_{12}	f_{23}	f_{34}	…	f_{nl}
边界长度比	R_{12}	R_{23}	R_{34}	…	R_{nl}

表 5.6　　**1 维成分长度**

成分	t_l	…	t_k	ic_l	…	ic_m	oc_l	…	oc_u
长度	L_{tl}	…	L_{tk}	L_{icl}	…	L_{icm}	L_{ocl}	…	L_{ocu}

表 5.5 中,成分关联区域指两个无孔的内部连通区域边界交集成分之间的两个区域的(最短)边界形成的小区域,如图 5.4 所示。$A_{12}, A_{23}, \cdots, A_{n1}$ 指与区域(1,2),(2,3),…,(n,1) 对应的面积。扁率$\left(\text{定义为}:f_{ij}=\dfrac{A_{ij}}{L_{ij}}\right)$和边界长度比$\left(\text{定义为}:R_{ij}=\dfrac{le_{ij}}{se_{ij}}\right)$被用来描述成分区域的形状和边界组成的差异。其中,$f_{ij}$ 为成分区域(i,j)的扁率,A_{ij} 和 L_{ij} 分别为成分区域(i,j)的面积和周长,R_{ij} 表示成分关联区域(i,j)的边界长度比,le_{ij} 和 se_{ij} 分别为组成这个成分关联区域的较长的和较短的边界长度。

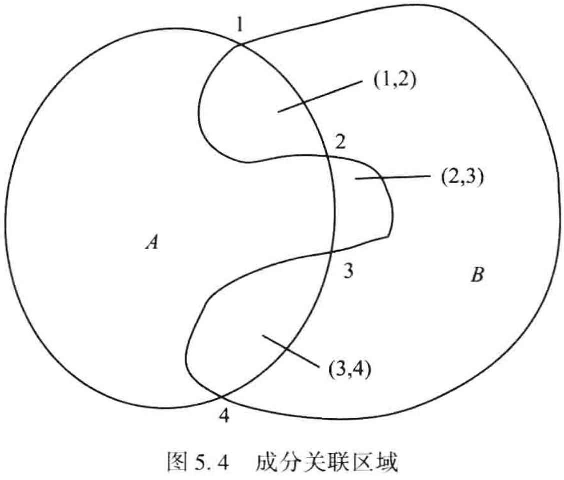


图 5.4　成分关联区域

表 5.6 中,1 维成分的度量,主要是 1 维相接成分的长度以及内部和外部交叉成分的长度。k,m 和 u 分别表示 1 维相接、内部交叉和外部交叉的数量,L 表示相应成分的长度,则 tk 表示第 k 个 1 维相接成分,icm 表示第 m 个 1 维内部交叉成分,ocu 表示第 u 个 1 维外部交

叉成分；L_{tk}，L_{icm} 和 L_{ocu} 分别表示其长度。

3）拓扑和度量相结合的空间关系描述

对于 2 个面目标，空间关系的名称、几何度量参数及其量值是描述空间关系的基本要素。当给定空间关系的名称 N，及其关于几何度量 m 的量值 v，则用有序三元组：

$$R = (N, m, v)$$

作为描述空间关系的基本单元。其中，量值 v 由 N 和 m 确定，记作 $v = m(N)$，则空间关系元表示为：

$$R = (N, m, m(N))$$

例如，两个简单区域 A 和 B 相离，其分离式接近性度量参数 DEB＝0.2，则该空间关系可描述为 R_{AB} ＝(DC，DEB，0.2)。

显然，此空间关系描述是将空间关系的定性描述和定量化描述放在一个统一的有序三元组中，能够较细致地描述空间关系，即不仅能区分“类间即为异”的关系，又能区分类内关系的差异。对于复杂区域的重叠关系，建立表 5.5 和表 5.6 进行描述。

3. 拓扑关系概念邻域和拓扑距离

(1)拓扑关系概念邻域

概念邻域源于 Allen(1983)的基于区间的定性时间推理，后来 Freksa(1991)把这一概念引入空间域中，Cohn 等人(1996)用概念邻域讨论了基于“蛋-黄”模型的不确定区域之间的 46 种拓扑关系。拓扑空间是建立在邻域概念上的，这里的邻域指的是位置 (x, y, z) 邻域。在 RCC 模型中，基本拓扑关系(JEPD 集中的关系)在概念上同样具有邻域，由于这一邻域主要针对拓扑关系概念而非点集拓扑中一个点的位置邻域，故称为概念邻域(conceptual neighbourhood)(高振记等，2008)。概念邻域关系具有某种相似的行为或性质，这对于描述和区分拓扑关系的组合以及拓扑关系的动态演化具有意义。

基于 RCC 模型的一个 JEPD 集中的任何两个关系，如果它们之间不经过任何第三个关系能够从一个关系连续、逐渐地过渡到另一个关系，则称这两个关系互为直接概念邻域，如 RCC-D-5 PP 和 PP^{-1} 为 EQ 的直接概念邻域。RCC-D-5 和 RCC-D-8 的直接概念邻域组合如图 5.5 和图 5.6 所示。

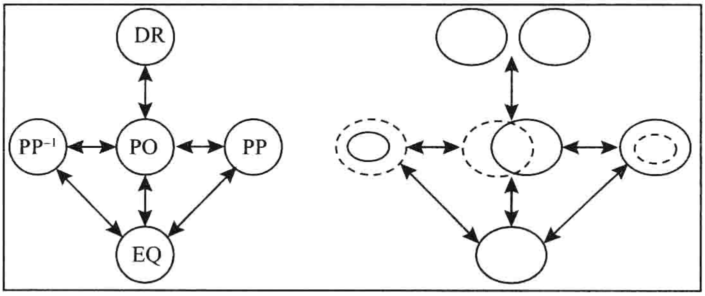


图 5.5 RCC-D-5 直接概念邻域图

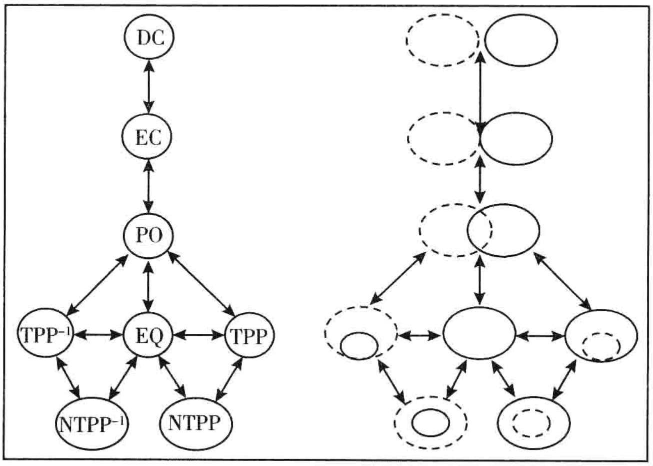
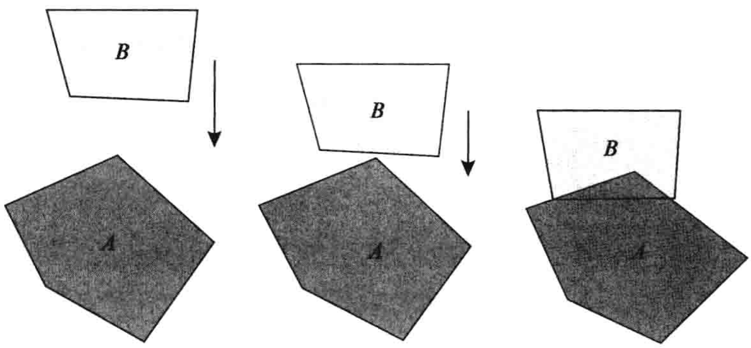


图 5.6 RCC-D-8 直接概念邻域图

基于 RCC 模型的一个 JEPD 集的任何一个子集,如果这个子集中的任何一对关系互为直接概念邻域,或由这个子集中的各元素组成的链中,任意邻接的一对关系互为直接概念邻域,则称这个子集为一个概念邻域。如果把图 5.5、图 5.6 看作一个无向图,RCC-D-5、RCC-D-8 的各级概念邻域可以看作是各个基本关系中路径长度相等的通路的集合,而 RCC-D-5、RCC-D-8 的概念邻域就是各级概念邻域的集合。高振记等人(2008)研究得出 RCC-D-5 共有 23 个概念邻域,RCC-D-8 的概念邻域为 97 个。

(2)拓扑距离

实际上,直接概念邻域表达了相邻的两个空间拓扑关系的直接转换过渡关系,而概念邻域则表达了任意一对空间拓扑关系之间的逐渐过渡关系,即反映了从一个拓扑关系连续转变到另一个拓扑关系的过程中需要经过的其他拓扑关系。空间目标间的拓扑关系是随目标位置(包括大小、形状、旋转、平移等)的变化而变化的。如图 5.7(a)所示,随着目标 *B* 的向下平移,拓扑关系依次变化为"邻接"、"重叠", *A* 和 *B* 之间的关系从"相离"变化到"重叠"必须经过"邻接"。这与概念邻域所表达的空间拓扑关系之间的逐渐过渡关系是一致的。这也在一定程度上说明了拓扑关系之间存在关联性,并且不同的拓扑关系之间的关联性可能不同,如"相离"与"邻接"间的关联性要比"相离"与"重叠"间的关联性强。这种关联性用拓扑距离来度量。



(a) 相离关系　　(b) 邻接关系　　(c) 重叠关系

图 5.7　拓扑关系变化的一种示例

在已有的关于拓扑关系的研究与实际应用中,多是基于 RCC-5 和 RCC-8 对应的 5 种和 8 种拓扑关系展开的,而且 4-交模型、9-交模型、4-交差模型推理出的简单区域间的拓扑关系与 RCC-8 的 8 种拓扑关系是一致的,所以关于拓扑距离的现有的研究成果中,均是探讨 8 种区域与区域关系之间的拓扑距离,邓敏等人(2005)定义了拓扑

距离函数，并依此计算了 8 种区域与区域之间关系的拓扑距离，吕秀琴等人(2006)将拓扑关系概念邻域图的节点看作是拓扑关系，通过计算节点之间的距离，得到了概念邻域图的差异矩阵。这里将他们的成果进行综合，并结合 RCC-D-8 直接概念邻域图，如图 5.5 所示，得到了 8 种拓扑关系之间的拓扑距离见表 5.7。

表 5.7　　**8 种区域与区域关系之间的拓扑距离**

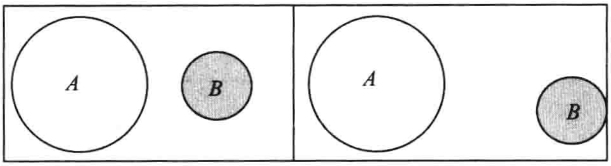
$D(-,-)$	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 $NTPP^{-1}$	内切于 TPP	包含于 NTPP	相等 EQ
相离 DC	0	1	2	3	4	3	4	3
邻接 EC		0	1	2	3	2	3	2
重叠 PO			0	1	2	1	2	1
被内切 TPP^{-1}				0	1	2	2	1
包含 $NTPP^{-1}$					0	2	2	1
内切于 TPP						0	1	1
包含于 NTPP							0	1

表 5.7 显示了 8 种区域与区域之间关系的拓扑距离，是对 8 种拓扑关系类型之间差异的一种量化方法。从表 5.7 中可以看出，该方法中的拓扑距离是离散的整数，同一种拓扑类型其拓扑距离是 0。事实上，即使同一种拓扑类型，相离的距离远近、相交面积的大小等是有差异的，而且在空间分析中这种差异是非常重要的，如图 5.8(a)比图 5.8(b)更接近邻接关系，它们与邻接关系的拓扑距离是不相等的，它们之间拓扑距离也不为 0。因此，这里用邓敏等人(2002、2009)、郭庆胜等人(2005)建立的区域与区域之间空间关系的量化描述(度量分析)方法，对拓扑距离进行扩展，其结果是所有拓扑关系之间的拓扑距离不再是离散的整数，而是连续的变化。式(5.5)和式(5.6)是扩展的拓扑距离的计算公式：

$$D(R_1,R_2) = m_1 + m_2 + d(R_1,R_2) - 1 \tag{5.5}$$

$$D(R_1,R_2) = m_2 - m_1 + d(R_1,R_2) \tag{5.6}$$

式(5.5)和式(5.6)中，R_1,R_2 均表示拓扑关系名称，$D(R_1,R_2)$ 表示任意两个拓扑关系之间的连续的拓扑距离，m_1 表示与拓扑关系 R_1 对应的几何度量参数，m_2 表示与拓扑关系 R_2 对应的集合度量参数，$d(R_1,R_2)$ 表示 R_1 与 R_2 之间离散的拓扑距离。



（a）相离1　　（b）相离2

图 5.8　空间实体相离的两种情形

如果 R_1 与 R_2 为同一种拓扑关系，则

$$D(R_1,R_2)=|m_1-m_2|$$

如图 5.8(a)与图 5.8(b)中 A 与 B 均为相离关系，他们之间扩展的拓扑距离为：

$$D(\mathrm{DC}_1,\mathrm{DC}_2)=|\mathrm{DEB}_1-\mathrm{DEB}_2|$$

如果 R_1 与 R_2 为不相同的拓扑关系，则结合表 5.7 与概念邻域图来决定用式(5.5)或式(5.6)，当 R_1 与 R_2 其中一个为相离关系时，用式(5.5)进行计算，其他情况用式(5.6)进行计算。将表 5.4 对应的度量参数代入式(5.5)或式(5.6)，得到对应的区域与区域之间关系扩展拓扑距离的计算公式见表 5.8，进行进一步计算，区域与区域之间关系扩展拓扑距离的取值范围见表 5.9。

表 5.8　**区域与区域之间关系扩展拓扑距离的计算公式**

$D(-,-)$	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 NTPP^{-1}	内切于 TPP	包含于 NTPP	相等 EQ
相离 DC	$\|\mathrm{DEB}_1-\mathrm{DEB}_2\|$	$\mathrm{DEB}_1+\mathrm{CB}_2$	$\mathrm{DEB}_1+\mathrm{CA}_2+1$	$\mathrm{DEB}_1+\mathrm{CB}_2+2$	$\mathrm{DEB}_1+\mathrm{CEC}_2+3$	$\mathrm{DEB}_1+\mathrm{CB}_2+2$	$\mathrm{DEB}_1+\mathrm{IC}_2+3$	DEB_1+3
邻接 EC	$\mathrm{CB}_1+\mathrm{DEB}_2$	$\|\mathrm{CB}_1-\mathrm{CB}_2\|$	$\mathrm{CA}_2-\mathrm{CB}_1+1$	$\mathrm{CB}_2-\mathrm{CB}_1+2$	$\mathrm{CEC}_2-\mathrm{CB}_1+3$	$\mathrm{CB}_2-\mathrm{CB}_1+2$	$\mathrm{IC}_2-\mathrm{CB}_1+3$	$3-\mathrm{CB}_1$

续表

$D(-,-)$	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 $NTPP^{-1}$	内切于 TPP	包含于 NTPP	相等 EQ
重叠 PO	CA_1+DEB_2+1	CA_1-CB_2+1	$\lvert CA_1-CA_2\rvert$	CB_2-CA_1+1	CEC_2-CA_1+2	CB_2-CA_1+1	IC_2-CA_1+2	$2-CA_1$
被内切 TPP_{-1}	CB_1+DEB_2+2	CB_1-CB_2+2	CB_1-CA_2+1	$\lvert CB_1-CB_2\rvert$	CEC_2-CB_1+1	CB_2-CB_1+2	IC_2-CB_1+2	$2-CB_1$
包含 $NTPP_{-1}$	CEC_1+DEB_2+3	CEC_1-CB_2+3	CEC_1-CA_2+2	CEC_1-CB_2+1	$\lvert CEC_1-CEC_2\rvert$	CB_2-CEC_1+2	IC_2-CEC_1+2	$2-CEC_1$
内切于 TPP	CB_1+DEB_2+2	CB_1-CB_2+2	CB_1-CA_2+1	CB_1-CB_2+2	CB_1-CEC_2+2	$\lvert CB_1-CB_2\rvert$	IC_2-CB_1+1	$2-CB_1$
包含于 NTPP	IC_1+DEB_2+3	IC_1-CB_2+3	IC_1-CA_2+2	IC_1-CB_2+2	IC_1-CEC_2+2	IC_1-CB_2+1	$\lvert IC_1-IC_2\rvert$	$2-IC_1$
相等 EQ	DEB_2+3	$3-CB_2$	$2-CA_2$	$2-CB_2$	$2-CEC_2$	$2-CB_2$	$2-IC_2$	0

表 5.9 **区域与区域之间关系扩展拓扑距离的取值范围**

$D(-,-)$	相离 DC	邻接 EC	重叠 PO	被内切 TPP^{-1}	包含 $NTPP^{-1}$	内切于 TPP	包含于 NTPP	相等 EQ
相离 DC	[0,1)	(0,2)	(1,3)	(2,4)	(3,5)	(2,4)	(3,5)	(3,4)
邻接 EC	(0,2)	[0,1)	(0,2)	(1,3)	(2,4)	(1,3)	(2,4)	(2,3)
重叠 PO	(1,3)	(0,2)	[0,1)	(0,2)	(1,3)	(0,2)	(1,3)	(1,2)
被内切 TPP^{-1}	(2,4)	(1,3)	(0,2)	[0,1)	(0,2)	(1,3)	(1,3)	(1,2)
包含 $NTPP^{-1}$	(3,5)	(2,4)	(1,3)	(0,2)	[0,1)	(1,3)	(1,3)	(1,2)
内切于 TPP	(2,4)	(1,3)	(0,2)	(1,3)	(1,3)	[0,1)	(0,2)	(1,2)
包含于 NTPP	(3,5)	(2,4)	(1,3)	(1,3)	(1,3)	(0,2)	[0,1)	(1,2)
相等 EQ	(3,4)	(2,3)	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)	0

表 5.8 中,"重叠"关系只用了一个度量参数(公共区域面积比、CA)进行计算,实际上,复杂区域的"重叠"关系的度量参数并不唯一。具体问题中,可先将关于重叠关系的所有度量参数都计算出来,

然后选取其中差异最大或者当时最为关心的参数进行比较，或者将多个度量参数合成为一个综合指标后计算。

通过分析比较，扩展的拓扑距离与离散的拓扑距离之间的差异在于：

①扩展的拓扑距离能够区分同一类拓扑关系之间的差异，而离散的拓扑距离不能；

②两类拓扑关系扩展的拓扑距离的最大值为 5，而最大的离散拓扑距离为 4；

③直接概念邻域的扩展的拓扑距离在区间(0,2)范围内取值，而其离散的拓扑距离为 1；

因此，扩展的拓扑距离在度量两个拓扑关系之间的距离时，比离散的拓扑距离更为准确。

5.2.2　不确定空间目标间拓扑关系表达模型

许多空间目标边界是不确定的。例如，野生动物的活动范围、山地和平原所在的区域等，因此，在空间数据库中经常需要考虑这些不确定空间目标之间的拓扑关系。而对于有确定边界的空间目标，因其位置数据存在不确定性，也会引起不确定拓扑关系。因此，不确定空间目标拓扑关系，一方面指模糊区域的拓扑关系，另一方面指空间数据位置不确定性引起的不确定拓扑关系。

在现有的不确定拓扑关系描述模型中，一种方法是获得两个不确定空间目标之间更多更精细的拓扑关系；另一种方法是基于人类思维对模糊概念的认识和理解，将不确定空间目标间精细的拓扑关系进行聚类或转化为模糊语言描述。

1. 不确定拓扑关系精细化表达模型

模糊区域之间拓扑关系表达模型均是在空间拓扑关系表达模型的基础上进行的扩展，如宽边界模型、“卵-黄”模型、4×4-交模型方法、模糊地理对象空间拓扑关系的粗集模型等。宽边界模型得出了 44 种拓扑关系、“卵-黄”模型得出了 46 种拓扑关系、4×4-交模型得出了 152 种拓扑关系、模糊地理对象空间拓扑关系的粗集模型得出了 505 种拓扑关系。这些模型的合理性都分别从数学或逻辑的角度

得到了证明,并且通过比较研究,这些模型在描述不确定空间目标之间的拓扑关系时是等价的。

(1)"卵-黄"模型

"卵-黄"模型("Egg-Yolk" Model)是 Cohn 等人(1994,1996)为了描述模糊区域之间的拓扑关系而建立的一种模型。在该模型中,一个模糊区域可以表示为一个卵(egg),该区域中确定的部分表示为一个卵黄(yolk),不确定的边界表示为卵白(white),这样就可以将一个模糊区域形象地表示为三个部分。"卵-黄"模型通过两个区域的卵和黄之间的 RCC-5 关系来对这两个模糊区域之间的拓扑关系进行分类。

设 A、B 是两个模糊区域,考察 A、B 的卵和黄之间的关系,可以建立完备和互斥的关系集合。这些关系可以用一个 2×2 矩阵来确定,矩阵中的每一个元素都是 RCC-5 关系中的一种,该矩阵为:

$$\begin{bmatrix} F(\mathrm{y}(A),\mathrm{y}(B)) & F(\mathrm{y}(A),\mathrm{e}(B)) \\ F(\mathrm{e}(A),\mathrm{y}(B)) & F(\mathrm{e}(A),\mathrm{e}(B)) \end{bmatrix}$$

其中,$\mathrm{y}(A)$ 是模糊区域 A 的黄,$\mathrm{e}(A)$ 是 A 的卵,$\mathrm{y}(B)$ 是模糊区域 B 的黄,$\mathrm{e}(B)$ 是 B 的卵;F 为二元空间 $R \times R$ 到 RCC-5 关系集合的一个映射。

由于矩阵的每一个元素都可能是 RCC-5 中的关系之一,因此该矩阵给出了 $5^4=625$ 种不同的可能关系。但是由于存在一些约束关系,如一个区域的卵黄必须包含于卵,并且每一个卵是非空的,因此最后可以得到 46 种不同的拓扑关系,如图 5.9 所示。

(2)宽边界区域 9-交集模型

根据 Egehnoefr 等人(1991)提出的 9-交集模型,Clmeentini 等人(1997)建立了宽边界区域之间的拓扑关系的表达模型。在这种模型中,宽边界区域 A 的内部(A^0)和宽边界(ΔA)分别对应于"卵-黄"模型中的卵黄和卵白,A 的外部(A^-)表示宽边界区域外部的部分。

考察下面的 3×3 矩阵,根据其中每个元素是空或者非空,可以确定两个宽边界区域 A 和 B 之间的 44 种不同的拓扑关系。

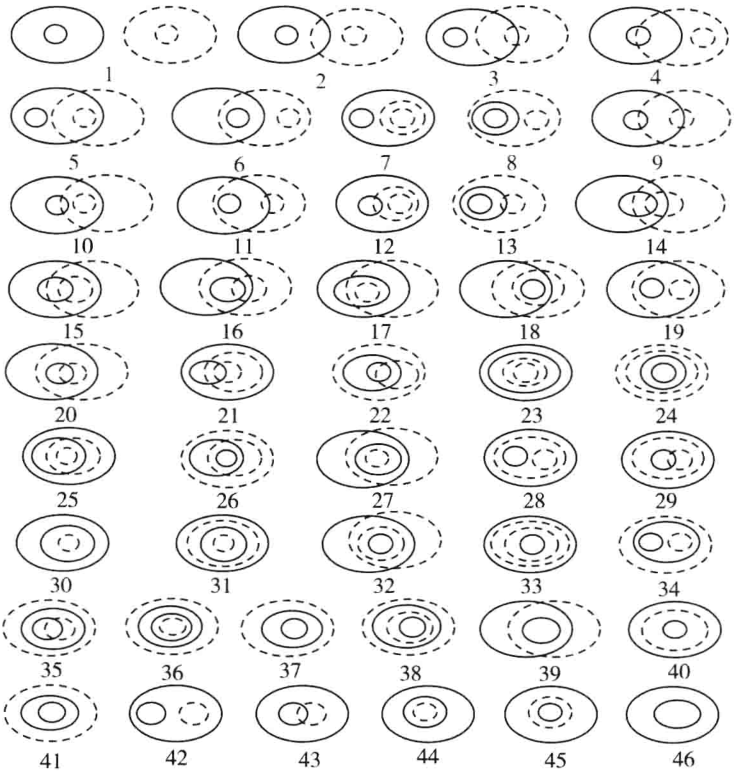


图5.9 基于RCC-5推理出的"卵-黄"拓扑关系图(Cohn,Gotts,1996)

$$\begin{bmatrix} A^0 \cap B^0 & A^0 \cap \Delta B & A^0 \cap B^- \\ \Delta A \cap B^0 & \Delta A \cap \Delta B & \Delta A \cap B^- \\ A^- \cap B^0 & A^- \cap \Delta B & A^- \cap B^- \end{bmatrix}$$

宽边界区域9-交集模型是基于点集理论方法,而"卵-黄"模型是基于逻辑表达和推理方法。这两种模型确定的44种关系和46种关系是一致的,只是在9-交集模型中不能区分"卵-黄"模型中的关系30与31、37与38,如图5.9所示,因此只能区分44种不同的拓扑关系。

(3)不确定拓扑关系的粗集模型

分别以 RCC-D-5 和 RCC-D-8 为基本拓扑关系集合，高振记等人(2008)研究了基于粗集的模糊地理对象间的拓扑关系。

RCC-D-5 模型中，两个模糊地理对象之间的拓扑关系可以用两个粗糙区域(各有一个上近似区域和下近似区域)的 4 个区域之间组成的矩阵表示，即

$$\begin{bmatrix} f(R_1^-, R_2^-) & f(R_1^-, R_{2-}) \\ f(R_{1-}, R_2^-) & f(R_{1-}, R_{2-}) \end{bmatrix}$$

其中，$f: R \times R \rightarrow \{\mathrm{DR}, \mathrm{PO}, \mathrm{PP}, \mathrm{PP}^{-1}, \mathrm{EQ}\}$ 为确定性区域(上下近似集)间的关系映射函数。为了能够用一种定量方法(而不仅仅是定性描述)对可能的拓扑关系进行描述，研究中引入相交、包含矩阵，用来表达模糊地理对象之间的任何一种可能的拓扑关系。每一种可能的拓扑关系都可以用一个上下近似集之间的相交和包含矩阵进行表达，该矩阵的形式表示如下：

$$\begin{bmatrix} R_1^- \cap R_2^- & R_1^- \cap R_{2-} & R_{1-} \cap R_2^- & R_{1-} \cap R_{2-} \\ R_1^- \subseteq R_2^- & R_1^- \subseteq R_{2-} & R_{1-} \subseteq R_2^- & R_{1-} \subseteq R_{2-} \\ R_2^- \subseteq R_1^- & R_{2-} \subseteq R_1^- & R_2^- \subseteq R_{1-} & R_{2-} \subseteq R_{1-} \end{bmatrix}$$

根据上述矩阵，得到确定对象之间的拓扑关系为 5 种，确定对象与模糊对象之间的拓扑关系为 22 种，模糊对象与模糊对象之间的拓扑关系为 46 种，最终基于 RCC-5 模型的离散空间的全部拓扑关系为 73 种。

RCC-D-8 模型中，模糊地理对象之间的空间拓扑关系研究要比基于 RCC-D-5 的研究复杂得多。对应于

$$\begin{bmatrix} f(R_1^-, R_2^-) & f(R_1^-, R_{2-}) \\ f(R_{1-}, R_2^-) & f(R_{1-}, R_{2-}) \end{bmatrix}$$

RCC-D-8 中确定性区域(上下近似集)之间的关系映射函数为：

$$f: R \times R \rightarrow \{\mathrm{DC}, \mathrm{EC}, \mathrm{PO}, \mathrm{TPP}, \mathrm{TPP}^{-1}, \mathrm{NTPP}, \mathrm{NTPP}^{-1}, \mathrm{EQ}\}$$

由于模糊对象与模糊对象组合之间的拓扑关系需要考虑模糊地理对象的上下近似区域为 TPP 和 NTPP 时的情景，这就使得 RCC-D-8的拓扑关系研究十分复杂。

对应地,由于基于 RCC-D-5 的模糊地理对象间空间拓扑关系的相交包含矩阵不能描述相接(EC)、正切真部分(TPP)和逆正切真部分(TPP^{-1})关系,因此,需要对相交包含矩阵进行扩充,进而表达基于 RCC-D-8 的模糊地理对象之间的拓扑关系。为此,引入相切和相邻的概念。相应地,全拓扑关系描述矩阵可以写为:

$$\begin{bmatrix} R_1^- \cap R_2^- & R_1^- \cap R_{2-} & R_{1-} \cap R_2^- & R_{1-} \cap R_{2-} \\ R_2^- \boxplus R_1^- & R_2^- \boxplus R_{1-} & R_{2-} \boxplus R_1^- & R_{2-} \boxplus R_{1-} \\ R_1^- \subseteq R_2^- & R_1^- \subseteq R_{2-} & R_{1-} \subseteq R_2^- & R_{1-} \subseteq R_{2-} \\ R_2^- \subseteq R_1^- & R_2^- \subseteq R_{1-} & R_{2-} \subseteq R_1^- & R_{2-} \subseteq R_{1-} \\ R_1^- \oplus R_2^- & R_1^- \oplus R_{2-} & R_{1-} \oplus R_2^- & R_{1-} \oplus R_{2-} \\ R_2^- \oplus R_1^- & R_2^- \oplus R_{1-} & R_{2-} \oplus R_1^- & R_{2-} \oplus R_{1-} \end{bmatrix}$$

离散空间中两个确定地理对象之间的拓扑关系,就是离散空间中 RCC-D-8 的 JEPD 集。RCC-D-8 的确定地理对象与模糊地理对象(C-R)组合中,当 $R_2^- = R_{2-}$ 时为 43 组拓扑关系;当 $R_1^- = R_{1-}$ 时也为 43 组拓扑关系。RCC-D-8 的模糊地理对象与模糊地理对象(R-R)组合描述十分复杂,通过概念邻域以及上下近似集之间的约束研究得出,离散空间中 R-R 组合基于 RCC-D-8 描述的空间拓扑关系共 505 种。

考虑到在实际应用中,拓扑关系往往由空间目标位置数据推理得到,邓敏等人(2007)以 4-交差模型为例,基于粗集的方法分析和推理了由空间位置数据不确定性引起的不确定拓扑关系。具体方法是首先计算目标 A 和 B 的粗集表达结果($\overline{A}$、$\underline{A}$、$\overline{B}$、$\underline{B}$),然后分别定义上下近似的内部和边界,最后依据 4-交差矩阵模型的形式化表达得到两个目标位置数据都带有不确定性的情况下,目标 A 和 B 之间的如下四个 4-交差矩阵:

$$R_{4ID}(\underline{A},\underline{B}) \quad R_{4ID}(\underline{A},\overline{B})$$

$$R_{4ID}(\overline{A},\underline{B}) \quad R_{4ID}(\overline{A},\overline{B})$$

利用以上四个4-交差矩阵进行计算,可以得到两个位置数据都带有不确定性的目标 A 和 B 之间的所有可能拓扑关系。

将具有位置不确定性的空间目标进行粗集表达,其结果与模糊地理目标粗集表达的形式是一致的。因此,高振记等人(2008)基于粗集研究的模糊地理对象间可能的拓扑关系同样适用于具有位置不确定性的空间目标间的拓扑关系表达。

2. 不确定拓扑关系的最可能识别

不确定拓扑关系精细化表达的结果导致不确定性地理对象间的拓扑关系数量众多,如基于 RCC-D-8 的模糊地理对象空间拓扑关系,可能的拓扑关系达505种之多。但是,人们一般不会去考虑模糊地理对象 A 与模糊地理对象 B 之间内边界与内边界、内边界与外边界、外边界与外边界是如何相交、相切、相离的。根据人们的常识性知识,最关心的一般是对象 A 与对象 B 是在一起(外接 EC)还是相离(DC)或相交(PO,TPP,NTPP)等。于是,史文中等人(2005)、何建华等人(2008)、邬伦等人(2010)基于人类认知的特点将不确定的拓扑关系的精细化表达演化到人们习惯接受的确定的拓扑关系类型中。

1)不确定性空间目标间的13种拓扑关系

史文中等人(2005)把任意一个不确定的空间目标 K 分成三个成分:中心区域 RK、外壳 BK 和外部 EK,基于两个目标的中心和外壳成分,将任意两目标 $K1$ 和 $K2$ 的拓扑关系表示为一个逻辑关系和13个基本关系类型:不相交关系、接触关系、几乎接触关系、几乎重叠关系、重叠关系、几乎被覆盖关系、几乎覆盖关系、被覆盖关系、覆盖关系、内部关系、包含关系、几乎相等关系、相等关系,这些关系分别定义如下:

①不相交关系

$$\langle K1,\text{不相交},K2\rangle \Leftrightarrow (K1 \cap K2 = \varnothing)$$

②接触关系

$$\langle K1,\text{接触},K2\rangle \Leftrightarrow (RK1 \cap RK2 = \varnothing) \wedge (BK1 \cap BK2 \neq \varnothing) \wedge (RK1 \cap RK2 = \varnothing) \wedge (BK1 \cap BK2 = \varnothing)$$

③几乎接触关系

$$\langle K1,\text{几乎接触},K2\rangle \Leftrightarrow (RK1 \cap RK2 = \varnothing) \wedge (BK1 \cap BK2$$

$\neq \varnothing) \wedge ((RK1 \cap RK2) \cup (BK1 \cap BK2) = \varnothing)$

④几乎重叠关系

$\langle K1,$几乎交叠$,K2\rangle \Leftrightarrow (RK1 \cap RK2 = \varnothing) \wedge (BK1 \cap BK2 \neq \varnothing) \wedge (RK1 \cap RK2 \neq \varnothing) \wedge (BK1 \cap BK2 \neq \varnothing)$

⑤重叠关系

$\langle K1,$交叠$,K2\rangle \Leftrightarrow (RK1 \cap RK2 \neq \varnothing) \wedge (RK1 \not\subset K2) \wedge (RK2 \not\subset K1)$

⑥几乎被覆盖关系

$\langle K1,$几乎被覆盖$,K2\rangle \Leftrightarrow (RK1 \subseteq K2)$

⑦几乎覆盖关系

$\langle K1,$几乎覆盖$,K2\rangle \Leftrightarrow (RK2 \subseteq K1)$

⑧被覆盖关系

$\langle K1,$被覆盖$,K2\rangle \Leftrightarrow (RK1 \subseteq RK2) \wedge (K1 \subseteq K2)$

⑨覆盖关系

$\langle K1,$覆盖$,K2\rangle \Leftrightarrow (RK1 \subseteq RK2) \wedge (K2 \subseteq K1)$

⑩内部关系

$\langle K1,$内部$,K2\rangle \Leftrightarrow (K1 \subseteq RK2)$

⑪包含关系

$\langle K1,$包含$,K2\rangle \Leftrightarrow (K2 \subseteq RK1)$

⑫几乎相等关系

$\langle K1,$几乎相等$,K2\rangle \Leftrightarrow (RK1 \subseteq K2) \wedge (RK1 \subseteq K1)$

⑬相等关系

$\langle K1,$相等$,K2\rangle \Leftrightarrow (K1 \subseteq K2) \wedge (K1 \subseteq K2)$

例如,对不确定的情形,二维空间中两个平面的拓扑关系如图5.10所示。

上面给出的拓扑关系是互斥的。也就是说,上面的任何两个$K1$和$K2$间的不同类型的关系不可能同时成立。另外,这些关系类型构成了所有可能拓扑关系的一个完备集。因此,这些关系类型可以作为拓扑关系集的一个分割。任意两个给定目标的拓扑关系一定属于这些关系类型中的某一个。

2)不确定拓扑关系聚类

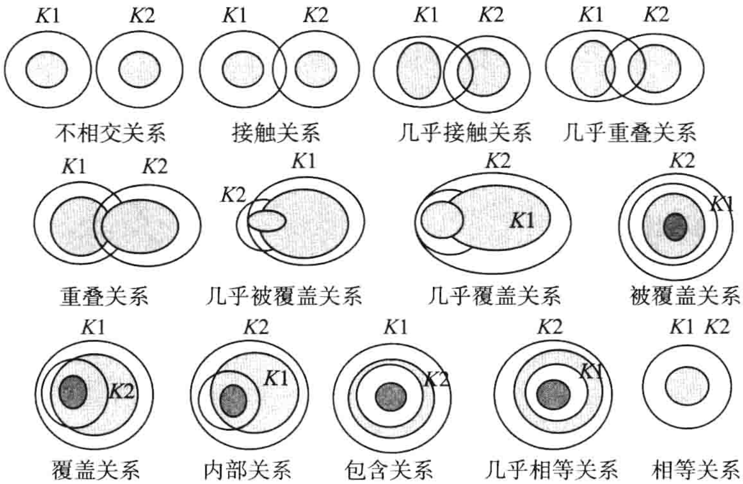


图 5.10 两个不确定性空间目标的拓扑关系

邬伦等人(2010)根据模糊地理对象之间拓扑关系的演化组合(即概念邻域组合)把离散空间中模糊地理对象之间的空间拓扑关系根据概念邻域进行分类,得出不同的拓扑关系组合,即拓扑关系聚类。

概念邻域反映了空间拓扑关系之间的演变关系,RCC 中的直接概念邻域表示 JEPD 集中的两个关系之间的连续过渡关系,概念邻域表示 JEPD 集中任意 n 个($2 < n \leqslant 8$) 关系之间的连续变化关系。将基于 RCC-D-5 和 RCC-D-8 得出的离散空间中模糊地理对象之间的空间拓扑关系根据概念邻域的作用范围相同和相近的概念邻域聚合为一类。具体的聚类原则为:

①直接概念邻域和概念邻域独立成类;

②不属于直接概念邻域和概念邻域的拓扑关系,如果其最近直接概念邻域或概念邻域存在,则划入其最近直接概念邻域或概念邻域;

③不属于①和②的拓扑关系向最近最小(概念邻域秩最小)的概念邻域聚合。

聚类要以直接概念邻域和最近概念邻域聚合为原则,即找出最具有拓扑相似性的“近亲”,当在最近一级上不能够进行聚类时,寻求高一级(远一级)的概念邻域进行聚类。

最终将基于 RCC-D-5 的模糊地理对象之间的 73 种拓扑关系聚类为 16 个概念邻域,具体的概念邻域种类及每个概念邻域包含的拓扑关系数目及拓扑编号见邬伦等人(2010)的专著《地理信息系统中的不确定性问题》。

基于 RCC-D-8 的模糊地理对象空间拓扑关系的聚类比较复杂。通过分析得出的 505 种空间拓扑关系共有 103 种互不重复的空间拓扑关系描述,根据以上的聚类原则,把基于 RCC-D-8 的 505 种模糊地理对象间的拓扑关系划分为 42 类,每一大类中的拓扑关系有着相同或近似的演化规律。互为直接概念邻域的两对拓扑关系可以直接互相演化,概念邻域级数越高,其中的拓扑关系之间的演化规律越不显著,或者说要通过很多的其他拓扑关系才能演变过来。

3)基于模糊贴近度分析的不确定拓扑关系表达模型

基于模糊贴近度确定不确定空间目标间最可能的拓扑关系是何建华等人(2008)提出的,具体方法遵照图 5. 11(模糊模式识别过程示意图),图 5. 11 中,模糊模式识别包括 3 个最基本组成要素,分别为:用于模式识别的模糊模式库、模式识别特征和模式匹配方法。因此,基于模糊贴近度分析的不确定空间目标最可能拓扑关系识别由

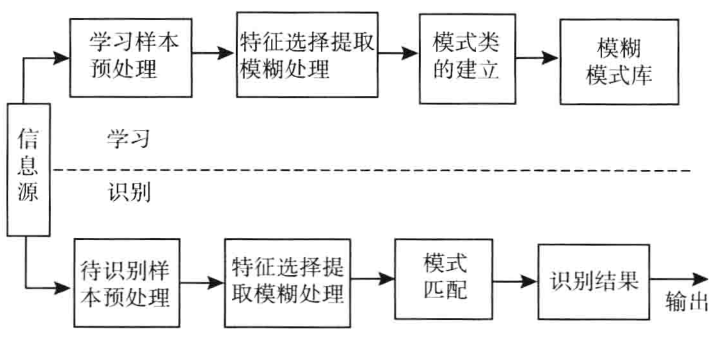


图 5. 11　模糊模式识别过程示意图

三步完成，一是拓扑关系模糊模式库的建立，二是拓扑关系度量特征选取，三是基于模糊贴近度的模式匹配方法。

(1)拓扑关系模糊模式库建立

这里的模糊模式库也就是基本拓扑关系概念集合，将 RCC-5 确定的拓扑关系谓词集合{DR, PO, PP, PP^{-1}, EQ}作为模糊模式库。

(2)拓扑关系度量特征选取

RCC 模型是以连接关系 C 为基本度量算子的，为方便利用现有的 GIS 分析功能实现空间关系推理运算和避免 C 算子的语义不确定性，以部分关系 P 代替连接算子 C，作为模型基本操作算子。部分关系 P 定义为：

$$P(X,Y) \equiv \text{def}[\ \forall x \in X \rightarrow x \in Y]$$

在操作算子 $P(X,Y)$ 的基础上，用三元组 $\{P(X,Y), P(X, \neg Y), P(Y,X)\}$ 来描述 RCC-5 模型的全部拓扑关系，具体见表 5.10（其中 $\neg Y$ 为 Y 的补，即 Y 的外部）。

表 5.10　**基于 $P(X,Y)$ 的 RCC-5 模型**

RCC-5 模型		P 三元组模型		
谓词	语义	$P(X,Y)$	$P(X,\neg Y)$	$P(Y,X)$
DR	相离	F	T	F
PO	相交	F	F	F
EQ	相等	T	F	T
PP	包含	T	F	F
PP^{-1}	包含于	F	F	T

依据表 5.10，拓扑关系可用一个三元组向量加以度量 $\text{Top} = (X,Y) = (P(X,Y), P(X,\neg Y), P(Y,X))$，由于基本的拓扑关系集合是五个确定的拓扑关系，且它对应的向量元素取值为 0 或 1，即 DR = (0,1,0)，PO = (0,0,0)，PP = (1,0,0)，PP^{-1} = (0,0,1)，EQ = (1,0,1)，则由 5 个基本拓扑关系谓词构成的基本模式库可表达为矩阵：

$$M\,\mathrm{top}_{5\times 3}=\begin{pmatrix}\mathrm{DR}\\ \mathrm{PO}\\ \mathrm{PP}\\ \mathrm{PP}^{-1}\\ \mathrm{EQ}\end{pmatrix}=\begin{bmatrix}P_{\mathrm{DR}}(X,Y) & P_{\mathrm{DR}}(X,\neg Y) & P_{\mathrm{DR}}(Y,X)\\ P_{\mathrm{PO}}(X,Y) & P_{\mathrm{PO}}(X,\neg Y) & P_{\mathrm{PO}}(Y,X)\\ P_{\mathrm{PP}}(X,Y) & P_{\mathrm{PP}}(X,\neg Y) & P_{\mathrm{PP}}(Y,X)\\ P_{\mathrm{PP}^{-1}}(X,Y) & P_{\mathrm{PP}^{-1}}(X,\neg Y) & P_{\mathrm{PP}^{-1}}(Y,X)\\ P_{\mathrm{EQ}}(X,Y) & P_{\mathrm{EQ}}(X,\neg Y) & P_{\mathrm{EQ}}(Y,X)\end{bmatrix}$$

$$=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\\ 1 & 0 & 1\end{pmatrix}$$

确定拓扑空间中对 P 的定义是基于二值逻辑的，在模糊拓扑空间中应用该关系还需要进行相应的拓展。正如确定拓扑空间到模糊拓扑空间可以通过从二值逻辑扩展为多值逻辑或模糊逻辑而实现对模糊拓扑空间目标的表达，相应地，P 算子也可以通过同样的方法进行扩展，即利用 $[0,1]$ 区间值对 P 进行定义：

$$P(X,Y)=\frac{A(X\cap Y)}{A(X)} \tag{5.7}$$

其中，$A(X\cap Y)$ 为不确定区域 X 与 Y 相交部分的模糊面积，$A(X)$ 为不确定区域 X 的模糊面积。由此定义可以看出，$P(X,Y)$ 的取值范围即为 $[0,1]$，当模糊集合 $X\subset Y$ 时，$P(Y,X)=1$，当模糊集合 $X\cap Y=\varnothing$ 时，$P(Y,X)=0$。

(3)基于模糊贴近度的模式匹配方法

基于拓扑关系模式库进行不确定空间目标之间空间关系实例与模式空间的匹配，进而求出不确定空间目标之间的拓扑关系分别隶属于5个基本拓扑关系的程度。在此，选择模糊贴近度作为模式匹配方法，模糊贴近度用于度量两模糊集合之间的距离，计算公式为：

$$N(X,Y)=1-\frac{1}{n}\sum_{k=1}^{n}\left|\mu_{X}(u_k)-\mu_{Y}(u_k)\right|^{\frac{1}{k}}$$

根据 k 的取值不同，该贴近度公式可分别称为海明(Hamming)贴近度、欧氏(Euclid)贴近度和明可夫斯基(Minkowski)贴近度等。在此取海明贴近度作为拓扑关系模式匹配的方法，即相应的计算公

式为：

$$N(\mathrm{top}_i,\mathrm{top}_{XY}) = 1 - \frac{1}{3}\sum_{j=1}^{3} | \mathrm{top}_{ij} - \mathrm{top}_{XYj} | \qquad (5.8)$$

这里，top_i 为 5 个基本拓扑关系之一，即 DR，PO，PP，PP^{-1} 和 EQ，top_{XY} 为不确定区域 X 和 Y 对应的可能的拓扑关系，top_{ij} 为第 i 个拓扑关系 P 三元组的第 j 个元组，top_{XYj} 则为不确定区域 X 和 Y 对应拓扑关系三元组的第 j 个分量，"‖"为两拓扑关系三元组间的海明距离。

由于 top_{ij} 的取值范围是 0 或 1，top_{XYj} 的取值范围为[0,1]，因此，$N(\mathrm{top}_i,\mathrm{top}_{XY})$ 的取值范围为[0,1]。由此可以看出，两拓扑关系三元组向量之间的贴近度越大，两空间拓扑关系概念越接近，因此就可以将贴近度定义为不确定空间目标间拓扑关系隶属于基本拓扑关系的隶属度。

(4)基于模糊贴近度分析的不确定拓扑关系表达步骤

根据图 5.11 所示的模糊模式识别的识别过程，可以看出基于模糊贴近度分析的不确定空间目标之间拓扑关系分析的步骤为：

第一步，根据不确定空间目标间的空间状态，根据式(5.7)分别计算拓扑关系三元组的向量 top_{XY} 各元组 $P(Y,X)$，$P(X,\neg Y)$ 和 $P(Y,X)$ 的值。

第二步，依据式(5.8)，计算向量 top_{XY} 与基本空间拓扑关系向量 top_i 之间的贴近度。

第三步，根据贴近度计算结果，以 top_{XY} 与各基本拓扑关系向量之间的贴近度作为不确定空间目标之间拓扑关系隶属于基本拓扑关系集成的隶属度，并求取其最大隶属度作为其不确定空间目标间最可能的拓扑关系。

例如，计算得到某不确定区域灌木林地 F 与农田 C 之间的拓扑关系三元组向量为 $\mathrm{top}_{CF} = (0.265, 0.464, 0.303)$。然后，可根据式(5.8)计算 top_{CF} 与 top_i 之间的贴近度，分别为：

$$N(\mathrm{DR},\mathrm{top}_{CF}) = 0.632$$

$$N(\mathrm{PO},\mathrm{top}_{CF}) = 0.656$$

$$N(\mathrm{PP},\mathrm{top}_{CF}) = 0.499$$

$$N(\mathrm{PP}^{-1},\mathrm{top}_{CF}) = 0.525$$

$$N(\mathrm{EQ}, \mathrm{top}_{CF}) = 0.368$$

因此,可得到本例提出的灌木林与农田之间的不确定拓扑关系,以一个拓扑关系谓词描述为 PO,若用 2 个拓扑关系谓词描述则为 DR,PO,若以 3 个拓扑谓词描述,则为 DR,PO 和PP^{-1}。

5.2.3　拓扑关系不确定性粗集表达模型

基于模糊贴近度的不确定拓扑关系的表达模型、不确定拓扑关系聚类是基于人类认知的特点将不确定拓扑关系的精细化表达演化到人们习惯接受的确定的拓扑关系类型中。不确定空间目标的 13 种拓扑关系是对基本拓扑关系类型的扩展，引入了人类所习惯的近似、模糊认知的因素。从这些可以看出，现有的不确定拓扑关系的研究成果均是在不确定描述中寻找确定的表达，不含有对不确定性程度的定量化描述。

由于拓扑关系是由空间数据计算出来的，空间数据存在不确定性，拓扑关系必然存在不确定性，根据误差传播定律，是可以计算出拓扑关系不确定性的大小的。但是由于拓扑关系计算的复杂性，以及拓扑关系是一种定性描述结果，用误差传播定律计算其不确定性是非常复杂的。然而，粗集是不确定性描述、推理、量化等的一种非常好的数学方法，已有学者将其应用于空间位置不确定性、空间关系不确定性的研究中。因此，这里以粗集为基础研究拓扑关系的不确定性问题。

该方法首先是建立含有位置不确定性的空间目标间的拓扑和度量相结合关系的粗集表达；然后计算上下近似间扩展的拓扑距离；最后用扩展的拓扑距离及其标准化形式作为不确定拓扑关系的量化指标。在这里仅研究区域与区域之间的拓扑关系。

1. 拓扑关系的粗集表达

在粗集空间中，模糊区域的粗集表达和空间目标位置不确定性的粗集表达的定义和形式是一致的，因此，模糊区域之间拓扑关系的粗集表达模型均可用于确定由于空间目标位置不确定性引起的不确定拓扑关系。设区域 A 为参照目标，区域 B 为空间查询和分析对象，A 和 B 的位置数据均含有不确定性,则 A 和 B 粗集表达分别为

$\overline{A}$ 、$\underline{A}$ 和 $\overline{B}$ 、$\underline{B}$ 。

“蛋-黄”模型、宽边界模型、4×4 模型都是对区域间所有可能拓扑关系的罗列，不区分这些关系孰轻孰重，也不区分这些关系是通过区域的内边界还是外边界形成的。这里在现有方法的基础上，首先将各种可能的拓扑关系是区域哪一部分形成的加以区分，然后计算与这些拓扑关系对应的度量参数的值，最后计算每两种拓扑关系之间的扩展的拓扑距离，取扩展的拓扑距离最大的两种拓扑关系分别为这两空间区域的拓扑关系的上近似和下近似。具体构建过程和方法如下：

①选择一种拓扑关系描述模型用于描述目标 A 和 B 之间的拓扑关系，这里选用4-交模型。4-交模型中，空间目标 A 与 B 之间的拓扑关系记作：

$$R_4(A,B)=\begin{pmatrix}\partial A\cap\partial B & \partial A\cap B^0\\ A^0\cap\partial B & A^0\cap B^0\end{pmatrix}$$

上式中，A^0 和 ∂A 分别为 A 的内部和边界，B^0 和 ∂B 分别为 B 的内部和边界。

相应地，$\overline{A}$ 、$\underline{A}$ 、$\overline{B}$ 、$\underline{B}$ 也分别分为内部和边界两部分，分别记作 $\overline{A}^0$ 和 $\partial\overline{A}$ 、$\underline{A}^0$ 和 $\partial\underline{A}$ 、$\overline{B}^0$ 和 $\partial\overline{B}$ 、$\underline{B}^0$ 和 $\partial\underline{B}$ 。$\overline{A}$ 、$\underline{A}$ 与 $\overline{B}$ 、$\underline{B}$ 之间的拓扑关系可用4 个 4–交模型来表示：

$$R_4(\overline{A},\overline{B})=\begin{pmatrix}\partial\overline{A}\cap\partial\overline{B} & \partial\overline{A}\cap\overline{B}^0\\ \overline{A}^0\cap\partial\overline{B} & \overline{A}^0\cap\overline{B}^0\end{pmatrix}$$

$$R_4(\overline{A},\underline{B})=\begin{pmatrix}\partial\overline{A}\cap\partial\underline{B} & \partial\overline{A}\cap\underline{B}^0\\ \overline{A}^0\cap\partial\underline{B} & \overline{A}^0\cap\underline{B}^0\end{pmatrix}$$

$$R_4(\underline{A},\overline{B})=\begin{pmatrix}\partial\underline{A}\cap\partial\overline{B} & \partial\underline{A}\cap\overline{B}^0\\ \underline{A}^0\cap\partial\overline{B} & \underline{A}^0\cap\overline{B}^0\end{pmatrix}$$

$$R_4(\underline{A},\underline{B})=\begin{pmatrix}\partial\underline{A}\cap\partial\underline{B} & \partial\underline{A}\cap\underline{B}^0\\ \underline{A}^0\cap\partial\underline{B} & \underline{A}^0\cap\underline{B}^0\end{pmatrix}$$

②确定与 $R_4(\overline{A},\overline{B})$ 、$R_4(\overline{A},\underline{B})$ 、$R_4(\underline{A},\overline{B})$ 、$R_4(\underline{A},\underline{B})$ 对应的度量参数 m_1 、m_2 、m_3 、m_4 ，并计算其值 $m_1(\overline{A},\overline{B})$ 、$m_2(\overline{A},\underline{B})$ 、$m_3(\underline{A},\overline{B})$ 、$m_4(\underline{A},\underline{B})$ 。

③确定 A 与 B 的拓扑和度量相结合的空间关系粗集表达。

含有位置不确定性的目标 A 与 B 的拓扑和度量相结合的空间关系表达为：

$$\begin{aligned}R(A,B)=\{&(R_4(\overline{A},\overline{B}),m_1,m_1(\overline{A},\overline{B}))、(R_4(\overline{A},\underline{B}),m_2,\\&m_2(\overline{A},\underline{B}))、(R_4(\underline{A},\overline{B}),m_3,m_3(\underline{A},\overline{B})),(R_4(\underline{A},\underline{B}),\\&m_4,m_4(\underline{A},\underline{B}))\}\end{aligned}\tag{5.9}$$

计算四种拓扑关系中每两种之间的扩展的拓扑距离，分别为 D_{12} 、D_{13} 、D_{14} 、D_{23} 、D_{24} 、D_{34} ，并计算这六个值中的最大值 $D_{\max}$ ，即

$$D_{\max}=\max\ (D_{12},D_{13},D_{14},D_{23},D_{24},D_{34})$$

若 $D_{st}=D_{\max}$($s=1$ 或 2 或 3 或 4，$t=1$ 或 2 或 3 或 4)，则选定式(5.9)中的第 s 个和第 t 个关系，并将它们与拓扑关系概念邻域图进行对照，取位于拓扑关系概念邻域图上方的关系为 A 与 B 拓扑关系的下近似 $\underline{R}(A,B)$ ，位于拓扑关系概念邻域图下方的关系为 A 与 B 拓扑关系的上近似 $\overline{R}(A,B)$ 。

事实上，由于 $R_4(\underline{A},\underline{B})$ 的确定性程度大于其他的拓扑关系，$R_4(\overline{A},\overline{B})$ 的确定性程度小于其他的拓扑关系，因而，区域 A 与 B 直接的拓扑关系的下粗近似集为 $R_4(\underline{A},\underline{B})$ 、上粗近似集为 $R_4(\overline{A},\overline{B})$ ，也就是 $\underline{R}(A,B)=R_4(\underline{A},\underline{B})$ ，$\overline{R}(A,B)=R_4(\overline{A},\overline{B})$ ，$D_{\max}=D_{14}$ 。

于是，A 与 B 的基于 4-交模型的拓扑和度量相结合的空间关系粗集表达为拓扑与度量相结合的上近似 $\overline{R}(A,B)$ 、下近似 $\underline{R}(A,B)$ 、扩展的拓扑距离 D 的三元组为：

$$R(A,B)=(\underline{R}(A,B),\overline{R}(A,B),D))=((R_4(\underline{A},\underline{B}),m_4,$$

$$m_4(\underline{A},\underline{B})),(R_4(\overline{A},\overline{B}),m_1,m_1(\overline{A},\overline{B})),D) \quad (5.10)$$

式(5.10)中，$D = D_{14}$。

该方法是基于粗集的4-交模型的拓扑与度量关系的表达与计算方法。该方法可以扩展到邬伦等人(2010)基于RCC-8的拓扑关系粗集表达的聚类结果，即首先把任意量目标粗集表达结果进行聚类，计算其聚类结果对应概念邻域的扩展的拓扑距离，用于量化其不确定性程度。

2. 拓扑关系不确定性量化指标

式(5.10)中D为上近似$\overline{R}(A,B)$与下近似$\underline{R}(A,B)$之间扩展的拓扑距离。D越大，表明拓扑关系$R(A,B)$的不确定性越大，但是D的取值范围在区间[0,5]内，而人类在对确定或不确定度的认知过程中，更习惯于用区间[0,1]之间的数值进行衡量。因此，将其转换为百分数用于评价拓扑关系不确定的程度，具体转换方法为：

$$k = \frac{D}{5} \times 100\%$$

k越大，表明拓扑关系$R(A,B)$的不确定性越大。

与此同时，可将A与B的拓扑和度量相结合的空间关系粗集表达为：

$$R(A,B) = (\underline{R}(A,B),\overline{R}(A,B),k)$$

5.2.4 多尺度空间拓扑关系不确定性建模

在多尺度表达中，空间目标的不确定性不断增大，拓扑关系保持相同或持续降低其复杂性和细节，因此，多尺度表达间拓扑关系的关联性成为人们关心的主要问题，多尺度表达中空间拓扑关系的一致性、等价性、相似性的研究成了人们关注多尺度表达拓扑关系的重要内容。依据“任意目标的拓扑以及任意目标之间的拓扑关系随着不同尺度层次，必须保持相同或持续降低其复杂性和细节”这一重要思想，Egenhofer等人(1994)提出一个评估多重表示拓扑一致性的框架，杜晓初等人(2007)建立了多尺度表达中空间拓扑关系等价性评价模型，吕秀琴等人(2006)给出了空间对象拓扑关系

相似性的定义以及在数字景观模型中拓扑关系相似性的度量方法。这些研究成果都是对不断抽象的目标计算的拓扑关系进行比较，来检验或评价多尺度表达中拓扑关系是否满足一致性、等价性或相似性的要求。而不是从空间目标位置不确定性出发，探讨多尺度空间拓扑关系的不确定性。在空间数据的多尺度表达中，随尺度的改变，空间数据的位置不确定性发生了变化，相应的拓扑关系即使满足一致性、等价性或相似性的要求，仍然存在不确定性。本研究针对空间目标位置不确定性的变化讨论拓扑关系的不确定性。

1. 多尺度空间拓扑关系不确定性评价概念模型

前面研究了空间目标及其多尺度表达的不确定性和空间拓扑关系的不确定性模型，以它们为基础，可以得出多尺度空间拓扑关系不确定性模型。具体可按图 5. 12 所示的多尺度空间拓扑关系不确定性评价流程图进行，拓扑关系流程图也可以拆分为以下六个步骤：

①确定空间区域 A 和 B ，分别计算两面目标的位置不确定性及其粗集表达结果；

②按要求将空间区域进行不同程度的化简，得到各简化尺度上空间区域 A_1 、B_1 ，A_2 、B_2 ，A_3 、B_3……

③计算各简化尺度上两空间区域的位置不确定性和粗集表达；

④将各尺度上空间区域的粗集表达结果代入 4-交模型，计算拓扑和度量相结合关系的粗集表达；

⑤分析各简化尺度上拓扑和度量相结合关系的粗集表达结果，评价各尺度上拓扑关系的不确定性；

⑥对比各尺度拓扑关系不确定性，总结拓扑关系随尺度变化的规律。

2. 多尺度空间拓扑关系不确定性表达模型

对于两个空间区域，能够实现的有 8 种空间拓扑关系，如果两个空间区域的表达尺度发生了变化，两个空间区域的空间数据就可能发生了变化，由空间数据计算的拓扑关系也可能发生变化。

这里研究两空间区域均化简后，它们之间拓扑关系的变化情况。

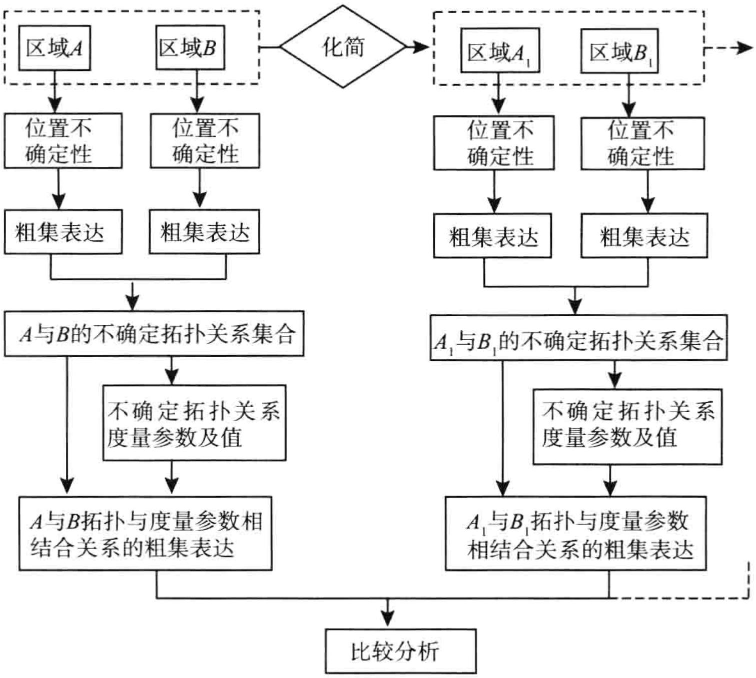


图 5.12 多尺度空间拓扑关系不确定性评价流程图

(1)多尺度空间对象拓扑关系的粗集表达

设多边形 A 具有 n 个顶点 $P_i=(x_i,y_i)$ 和 n 条边 $P_1P_2,P_2P_3,\cdots,P_{n-1}P_n$ 以及 P_nP_1，单个顶点的方差-协方差矩阵具有如下形式：

$$\sum_{P(i)P(i)}=\begin{bmatrix}\sigma_{x(i)}^2 & \sigma_{x(i)y(i)}\\ \sigma_{x(i)y(i)} & \sigma_{y(i)}^2\end{bmatrix}$$

多边形 B 具有 m 个顶点 $Q_j=(x_j,y_j)$ 和 m 条边 $Q_1Q_2,Q_2Q_3\cdots,Q_{m-1}Q_m$ 以及 Q_mQ_1，单个顶点的方差-协方差矩阵具有如下形式：

$$\sum_{Q(j)Q(j)}=\begin{bmatrix}\sigma_{x(j)}^2 & \sigma_{x(j)y(j)}\\ \sigma_{x(j)y(j)} & \sigma_{y(j)}^2\end{bmatrix}$$

利用 A 和 B 的顶点方差，根据 5.1 节的方法，可分别求得 $\overline{A}$ 、$\underline{A}$，

$\overline{B}$、$\underline{B}$。

将 A 和 B 进行不同程度的化简，若简化后的尺度为 s_i（$i=1,2,\cdots,n$），尺度 s_i 上区域为 A_i,B_i，根据 4.2.3 节的方法，可求得 A_i，B_i 的综合不确定性分别为 σ_{Ai}、σ_{Bi}，根据本章 5.1 节的方法，进一步求得 A_i,B_i 的粗集表达分别为 $\overline{A_i}$、$\underline{A_i}$，$\overline{B_i}$、$\underline{B_i}$。

利用 5.2.3 节中的拓扑关系不确定性的粗集表达方法，可得到初始尺度上区域 A 和 B 的拓扑关系粗集表达结果为：

$$R(A,B)=(\underline{R}(A,B),\overline{R}(A,B),D)$$

尺度 s_i 上区域 A_i 和B_i 的拓扑关系的粗集表达结果为：

$$R(A_i,B_i)=(\underline{R}(A_i,B_i),\overline{R}(A_i,B_i),D_i)$$

(2)多尺度空间对象拓扑关系不确定性度量

5.2.4 节中定义了各尺度上拓扑关系的上近似、下近似，以及上下近似之间的扩展的拓扑距离。每一尺度上扩展的拓扑距离可用来度量对应尺度上区域间拓扑关系的不确定程度。这里将各尺度上扩展的拓扑距离与初始尺度上扩展的拓扑距离进行比较，用简化尺度拓扑关系扩展的拓扑距离 D_i 与初始尺度扩展的拓扑距离 D 的差值 $\Delta D_i(i=1,2,\cdots,n)$表示简化尺度上拓扑关系相对于初始尺度拓扑关系的不确定性程度，用公式表达为

$$\Delta D_i=D_i-D$$

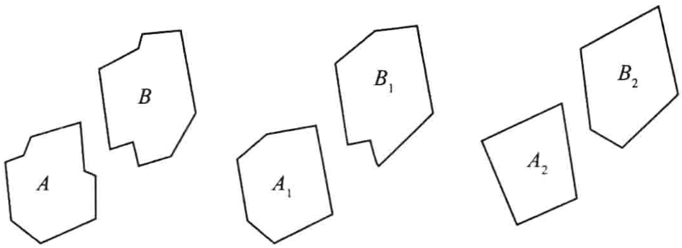
3. 实验与讨论

本实验选用的数据为图 5.13(a)所示的多边形 A 和 B，并将其进行两级尺度的化简，如图 5.13(b)、(c)所示，为了便于比较，将简化尺度图形扩展至原始图形大小。

根据 4.2.3 节中的方法计算原始数据及其简化尺度的综合不确定度，并用各尺度两倍的综合不确定度作为相应尺度的不确定带宽度，不确定带的内边界为多边形区域的下近似、外边界为多边形区域的上近似，如图 5.14 所示。

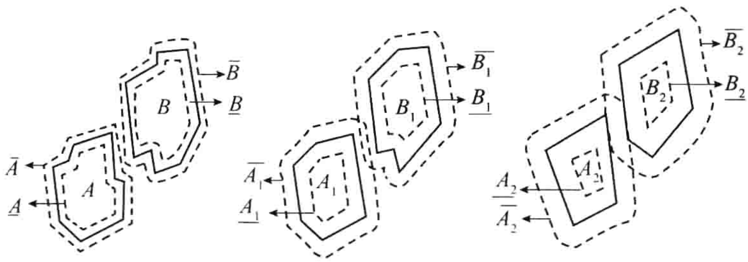
用 4-交模型来描述各尺度多边形之间的拓扑关系，得到

$$R_4(A,B)=\{R_4(\overline{A},\overline{B})、R_4(\overline{A},\underline{B})、R_4(\underline{A},\overline{B})、R_4(\underline{A},\underline{B})\}$$



（a）原始数据　　（b）尺度s_1　　（c）尺度s_2

图 5.13　实验数据



（a）原始数据　　（b）尺度s_1　　（c）尺度s_2

图 5.14　原始数据及其简化尺度的粗集表达

$$= \{\mathrm{DC}、\mathrm{DC}、\mathrm{DC}、\mathrm{DC}\}$$

$$R_4(A_1,B_1) = \{R_4(\overline{A_1},\overline{B_1})、R_4(\overline{A_1},\underline{B_1})、R_4(\underline{A_1},\overline{B_1})、R_4(\underline{A_1},\underline{B_1})\}$$
$$= \{\mathrm{PO}、\mathrm{DC}、\mathrm{DC}、\mathrm{DC}\}$$

$$R_4(A_2,B_2) = \{R_4(\overline{A_2},\overline{B_2})、R_4(\overline{A_2},\underline{B_2})、R_4(\underline{A_2},\overline{B_2})、R_4(\underline{A_2},\underline{B_2})\}$$
$$= \{\mathrm{PO}、\mathrm{DC}、\mathrm{DC}、\mathrm{DC}\}$$

与拓扑关系 DC 对应的度量参数为 DEB，与拓扑关系 PO 对应的度量参数为 CA，于是各尺度上拓扑和度量相结合的空间关系表达为：

$$R(A,B) = \{(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\overline{A},\overline{B})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\overline{A},\underline{B})),$$
$$(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A},\overline{B})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A},\underline{B}))\} \tag{5.11}$$

$$R(A_1,B_1) = \{(\mathrm{PO},\mathrm{CA},\mathrm{CA}(\overline{A_1},\overline{B_1})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\overline{A_1},\underline{B_1})),$$
$$(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A_1},\overline{B_1})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A_1},\underline{B_1}))\} \quad (5.12)$$

$$R(A_2,B_2) = \{(\mathrm{PO},\mathrm{CA},\mathrm{CA}(\overline{A_2},\overline{B_2})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\overline{A_2},\underline{B_2})),$$
$$(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A_2},\overline{B_2})),(\mathrm{DC},\mathrm{DEB},\mathrm{DEB}(\underline{A_2},\underline{B_2}))\} \quad (5.13)$$

将图 5.13(a)的数据代入式(5.11)进行计算得到：

$$R(A,B) = \{(\mathrm{DC},\mathrm{DEB},0.13),(\mathrm{DC},\mathrm{DEB},0.66),(\mathrm{DC},\mathrm{DEB},0.52),(\mathrm{DC},\mathrm{DEB},0.81)\}$$

$\rightarrow D_{12} = 0.53, D_{13} = 0.39, D_{14} = 0.68, D_{23} = 0.14, D_{24} = 0.15, D_{34} = 0.29$;

$\rightarrow D_{max} = D_{14} = 0.68$;

$\rightarrow R(A,B) = (\underline{R}(A,B),\overline{R}(A,B),D) = ((\mathrm{DC},\mathrm{DEB},0.81),(\mathrm{DC},\mathrm{DEB},0.13),0.68)$

将图 5.14(b)的数据代入式(5.12)进行计算得到：

$$R(A_1,B_1) = \{(\mathrm{PO},\mathrm{CA},0.03),(\mathrm{DC},\mathrm{DEB},0.71),(\mathrm{DC},\mathrm{DEB},0.45),(\mathrm{DC},\mathrm{DEB},0.88)\}$$

$\rightarrow D_1^{12} = 0.74, D_1^{13} = 0.48, D_1^{14} = 0.91, D_1^{23} = 0.26, D_1^{24} = 0.17, D_1^{34} = 0.43$;

$\rightarrow D_{max} = D_1^{14} = 0.91$;

$\rightarrow R(A_1,B_1) = (\underline{R}(A_1,B_1),\overline{R}(A_1,B_1),D_1) = ((\mathrm{DC},\mathrm{DEB},0.88),(\mathrm{PO},\mathrm{CA},0.03),0.91)$

将图 5.13(c)的数据代入式(5.13)进行计算得到：

$$R(A_2,B_2) = \{(\mathrm{PO},\mathrm{CA},0.09),(\mathrm{DC},\mathrm{DEB},0.60),(\mathrm{DC},\mathrm{DEB},0.41),(\mathrm{DC},\mathrm{DEB},0.89)\}$$

$\rightarrow D_2^{12} = 0.69, D_2^{13} = 0.50, D_2^{14} = 0.98, D_2^{23} = 0.19, D_2^{24} = 0.29, D_2^{34} = 0.48$;

$\rightarrow D_{max} = D_2^{14} = 0.98$;

$\rightarrow R(A_2,B_2) = (\underline{R}(A_2,B_2),\overline{R}(A_2,B_2),D_2) = ((\mathrm{DC},\mathrm{DEB},0.89),(\mathrm{PO},\mathrm{CA},0.09),0.98)$

也就是说,图 5.14(a)中,区域 A 与 B 拓扑关系的下近似为相离关系,分离式接近性度量参数 DEB=0.81;上近似为相离关系,分离式接近性度量参数 DEB=0.13;上下近似之间扩展的拓扑距离 D = 0.68。

图 5.14(b)中,区域 A_1 与 B_1 拓扑关系的下近似为相离关系,分离式接近性度量参数 DEB=0.88;上近似为重叠关系,公共区域面积比 CA=0.03;上下近似之间扩展的拓扑距离 D_1 = 0.91。

图 5.14(c)中,区域 A_2 与 B_2 拓扑关系的下近似为相离关系,分离式接近性度量参数 DEB=0.89;上近似为重叠关系,公共区域面积比 CA=0.09;上下近似之间扩展的拓扑距离 D_2 = 0.98。

将图 5.14(b)、图 5.14(c)扩展的拓扑距离与图 5.14(a)扩展的拓扑距离进行比较,得到:

$$\Delta D_1 = 0.91 - 0.68 = 0.23$$

$$\Delta D_2 = 0.98 - 0.68 = 0.30$$

$\Delta D_2 > \Delta D_1$,也就是说,随着多边形的化简,多边形之间拓扑关系的不确定程度增大。

5.3 多尺度空间方向关系不确定性建模

空间方向关系(闫浩文, 2001)存在于地理目标的两目标之间,是在一定的方向参考系统中从一个空间目标到另一个空间目标的指向, 通常用角度或东、南、西、北等术语表示。将参考目标、源目标和参照系统作为描述空间方向关系的三要素。在已有的空间方向关系描述模型中, 对空间方向区域进行硬性划分, 而且在计算空间目标的方向关系时, 不考虑空间目标的模糊性或空间数据的不确定性, 最终导出的两目标间的方向关系是明确的。但是, 空间方向片的跳跃式划分以及不考虑空间目标的模糊性或空间数据的不确定性, 都是不符合人类的认知逻辑的。因此, 基于已有的确定性空间方向关系模型, 有必要进一步定义符合人类认知逻辑的不确定方向关系模型或对空间方向关系的不确定性进行评价。

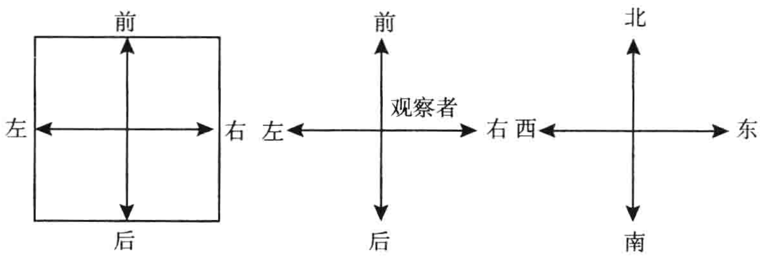
对于空间方向关系不确定性的研究, 一是定义平滑、连续的方

向概念，计算两空间目标对不同方向概念的隶属程度，这方面的研究还处于起步阶段，何建华(2006)研究了不确定空间目标方向关系的模糊描述模型；二是在已有方向关系模型中基于空间目标位置不确定性的空间方向关系的不确定性表达，这方面的研究主要使用粗集来表达，因为空间目标位置不确定性引起的空间方向关系的不确定性，主要的研究成果如杜世宏(2003)基于方向关系矩阵模型的空间方向关系粗集表达，作者进一步研究了基于粗集的空间方向关系不确定性的定量评价及多尺度空间方向关系的不确定性模型。

5.3.1　空间方向关系模型与空间方向关系的计算

1. 空间方向参考框架

空间方向的参考框架是由于不同目的、用途和观点对空间进行的划分。如果参考框架改变了，空间方向一般也会改变。Retz Schmidt(1988)将参考框架(也称为方向参考系统)分为3类，如图5.15(闫浩文，2001)所示：



(a)内部参考框架　(b)直接参考框架　(c)外部参考框架

图5.15　空间方向的参考框架

①内部参考框架(intrinsic reference frame)：该框架指一个对象在自身内部建立的方向参照系，多用前、后、左、右等词语描述，如图5.15(a)所示。例如，一栋居民住宅，正门入口为前，阳台为后，卧室居左，其他房间、物品的空间方向可以此确定。

②直接参考框架(deictic reference frame)：该框架基于观察者

的观点建立，一个观察者以自己的前、后、左、右等对空间进行划分，建立空间方向的参照系，如图5.15(b)所示。

③外部参照系(extrinsic reference frame)：地球表面上的外部参考框架可通过选择不同的北方向(磁北、真北、坐标北等)来建立，如图5.15(c)所示。相应的外部参考框架也有所区别。

内部参考框架一般在机器人导航、人工智能等领域得到较广泛的应用，它一般适用于小尺度空间，如一间房内的桌、椅、板凳等在房间内的方位关系。直接参考框架是一种自然语言表达的方向关系参照系统，不适于在数据库环境下分析和处理，但在研究自然语言的理解、智能导航等领域，要求进行方便的人机交互，则需要考虑直接参考框架与数据库环境下的内部参考框架或外部参考框架的转换的问题。外部参考框架是适用于研究大尺度范围空间目标方向关系的表达，并且易于在数字环境下进行建模和处理，是适用于GIS问题研究的方向关系参照系统。

2. 空间方向关系模型

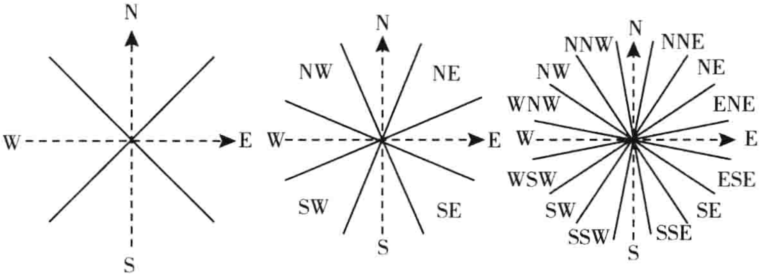
(1)锥形模型

锥形模型类似于解析几何的直角坐标系把平面空间分解为4个象限一样，从空间参考点出发将其周围的区域划分成若干具有方向性的子区域，并根据源目标与这些子区域相交的情况来确定源目标与其空间参考目标(点)之间的方向关系，常用的方向划分粒度为4方向、8方向、16方向，甚至粒度划分更细，如图5.16所示。

在四方向模型中，定义了四个方向关系：East(E)、North(N)、West(W)、South(S)；在八方向模型中，定义了八个方向关系：E、NE、N、NW、W、SW、S、SE；在十六方向模型中，定义了十六个方向关系：N、NNE、NE、ENE、E、ESE、SE、SSE、S、SSW、SW、WSW、W、WNW、NW、NNW。我们称这些方向为主方向，如果源目标落在某一方向的锥形区域内，则认为源目标在参考目标的该主方向上。

(2)最小外切矩形模型

最小外切矩形模型的基本思想是：先建立空间目标的最小外切矩形(MBR)，然后分别投影到横(纵)坐标轴上，两个空间目标在



（a）4方向系统　（b）8方向系统　（c）16方向系统

图 5.16　空间方向关系锥形模型

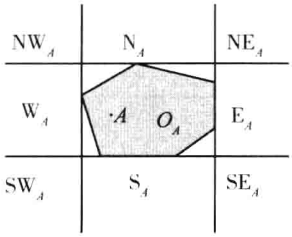
每个坐标轴上的投影会出现 Allen 的时间段关系，两个坐标轴上的这些关系的组合就表达了这两个空间目标之间的大致空间方向关系。该模型常用于面状目标之间空间方向关系的表达，也属于描述空间目标方向关系的投影模型。

(3)空间方向关系矩阵模型

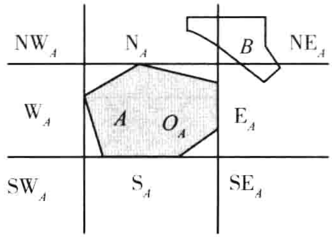
当参考空间目标为一个面时，就需要使用图 5.17 所示的方法描述空间方向关系，它也可以用一个 3×3 矩阵来表达，Goyal (2000)称此矩阵为"方向关系矩阵"(direction-relation matrix)。方向关系矩阵模型的基本思想是：如果参考空间目标是多边形，先建立其最小外切矩形(MBR)然后以 MBR 为中心把平面划分为 9 个区域(称为 9 个方向片)，每一个方向片对应一个主方向，如图 5.17(a)所示，参考空间目标所在的方向片为"Same"方向(O_A)，相对于参考空间目标 A 的方向集为 $\{N_A, NE_A, E_A, SE_A, S_A, SW_A, W_A, NW_A, O_A\}$，源目标 B 总会落到至少 1 个方向片中，如图 5.17(b)所示对 9 个方向片和源目标分别求交就可以构建方向关系矩阵，空间方向关系矩阵的一般形式为：

$$\mathrm{Dir}(A, B)=\begin{bmatrix} NW_A \cap B & N_A \cap B & NE_A \cap B \\ W_A \cap B & O_A \cap B & E_A \cap B \\ SW_A \cap B & S_A \cap B & SE_A \cap B \end{bmatrix}$$

矩阵元素的值不会全部为空，若交集为空，则值为 0，否则为



(a)



(b)

图 5.17 投影模型

1。图 5.17(b)所示的目标 A 和 B 之间的空间方向关系可以表示为：$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$。当目标 B 与不同的方向片相交的面积不一样时，Goyal(2000)用源目标在某一方向片上的面积与总面积的比值作为矩阵元素的值，改进后的矩阵形式如下：

$$\begin{pmatrix} \frac{\mathrm{Area}(\mathrm{NW}_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(\mathrm{N}_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(\mathrm{NE}_A \cap B)}{\mathrm{Area}(B)} \\ \frac{\mathrm{Area}(\mathrm{W}_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(O_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(\mathrm{E}_A \cap B)}{\mathrm{Area}(B)} \\ \frac{\mathrm{Area}(\mathrm{SW}_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(\mathrm{S}_A \cap B)}{\mathrm{Area}(B)} & \frac{\mathrm{Area}(\mathrm{SE}_A \cap B)}{\mathrm{Area}(B)} \end{pmatrix}$$

其中，Area 为面积函数，该矩阵的所有元素值之和为 1。

(4)基于方向 Voronoi 图(DVDs)的方向群模型

基于 DVDs 的方向群模型是 Haowen Yan 等人(2006)在基于 Voronoi 图方向关系模型的基础上，计算各 DVD 边对应方向的权重，DVD 边的方位角落在 4 方向系统或者 8 方向系统哪个方向区域内，源目标就部分位于参考目标的该主方向上，且其权重为方位角位于该主方向方位角范围内所有 DVD 边权重的和。源目标相对于参考目标的方向群及其权重在 4 方向系统中的表示见表 5.11。

表 5.11　　**基于 DVDs 的 4 方向系统方向关系描述**

主方向	重叠(Same)	北(N)	东(E)	南(S)	西(W)
各方向权重(%)	W_{Same}	W_N	W_E	W_S	W_W

表中 $W_{Same}+W_N+W_E+W_S+W_W=100$。

基于方向 Voronoi 图(DVDs)的方向群模型，是以 Gestalt 原则作为多方向关系认知的基础(与空间认知相吻合)，在方向关系判断之前先将源目标和参考目标进行形状概括(充分考虑了图形化简的需要，与空间目标多尺度空间关系的表达相一致，具有多尺度包容性)，然后计算 DVDs，再构建定量的方向群，最后构建 4 方向或者 8 方向 2 维表。

图 5.18 中参考目标 A、源目标 B 的基于 DVDs 的 4 方向系统方向关系描述见表 5.12。

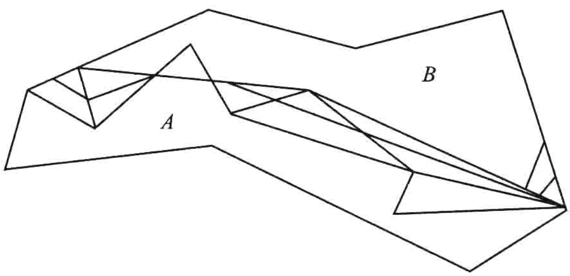


图 5.18　DVDs 的构建

表 5.12　**图 5.18 中基于 DVDs 的 4 方向系统方向关系描述**

主方向	重叠(Same)	北(N)	东(E)	南(S)	西(W)
各方向权重(%)	4	79	9	0	8

本方向模型中 DVDs 的建立是关键，依据基于方向 Voronoi 图(DVDs)的方向群模型的构建原理和过程，在计算 Dir_{AB} 时，Dir_{BA} 可同时计算出来，这也为空间推理和判断提供了便利。

3. 空间方向关系的计算

方向本身就是一种指向，这种指向可以用一个向量来描述，表示一个点相对另外一个点的方位或方向，并且可以准确地计算出来。当以点为参考时，它与其他源目标之间的空间方向关系可以使用锥形模型来描述，也可以进行定量计算，如图5.19(a)所示，点A与点B的空间方向关系可以用向量$\overrightarrow{AB}$与正北方向的夹角(顺时针)来描述。点A与直线段BC之间的空间方向关系可以用向量$\overrightarrow{AB}$和$\overrightarrow{AC}$分别与正北方向的夹角θ_1和θ_2表示，(θ_1，θ_2)是一个区间，也可以抽象为θ_1和θ_2的平均值。但是，同样一个(θ_1，θ_2)或一条角平分线的方位θ会代表差别比较大的两条不同的直线段的方向，如图5.19(b)所示。为了能区别直线段BC和DE分别与点A之间的方向关系，可以分别用L_2和L_1的方位表示，其中，F_1和F_2分别为DE和BC的中点，这里的L_1和L_2不是角平分线，而是A与直线段中点的连线。当然，也可以对直线段进行等分，用这些等分后的小直线段与点A的方向关系的平均值描述。

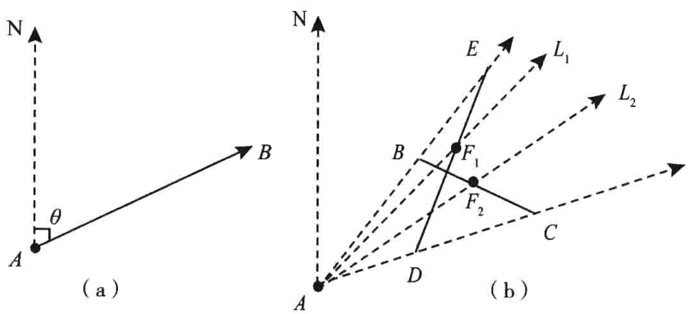


图5.19 基本空间方向关系的计算

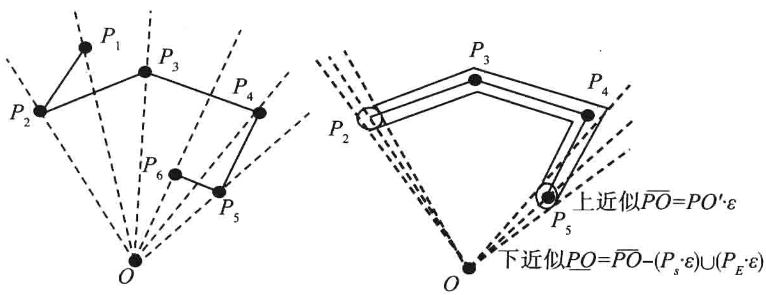
5.3.2 空间方向关系不确定性模型

1. 锥形模型中空间方向关系的粗集表达

(1)锥形模型中源目标的粗集表达

在GIS中，将确定性地理目标分为点目标、线目标、面目标。

已有确定性地理目标位置不确定性的粗集表达中，目标边界的不确定性即为粗集的边界。因此，为将源目标的粗集表达进一步用于定义不确定方向关系，应首先确定源目标在锥形模型中定义方向分布范围的边界。在锥形模型中，点的方向分布范围由点的位置确定，面的方向分布范围由面的边界线确定，线的方向分布范围比较复杂，如图 5.20(a)中的折线，其方向分布与 P_2、P_3、P_4、P_5 的连线是一致的，P_2、P_5 是定义线方向分布范围的边界点。



（a）折线　　　（b）折线方向关系的粗集表达

图 5.20　锥形模型中折线方向及其粗集表达

因此，当源目标 PO 为线时(由节点 P_1、$P_2 \cdots P_n$ 及其连线构成)，首先寻找折线在锥形模型中方向分布的边界点，边界点是以参考点为视点的线目标的可视部分的外边界点。边界点的寻找过程是：以参考目标(点)为坐标原点、任一从坐标原点发出的与线目标不相交的射线方向为极轴建立极坐标系，线目标 PO 上节点 P_i $(i=1,2,\cdots,n)$ 的极坐标为 $P_i(\alpha_i,\rho_i)$，计算 $\alpha_{max}=\max(\alpha_1,\alpha_2,\cdots,\alpha_n)$，$\alpha_{min}=\min(\alpha_1,\alpha_2,\cdots,\alpha_n)$，在线目标 PO 上节点中查找其极角与 α_{max} 相等的节点 P_S 和与 α_{min} 相等的节点 P_E，若 P_S 和 P_E 不止一个，分别取其中极径最小的节点为 P_S 和 P_E，然后在线目标上截下以 P_S 和 P_E 为端点的折线段即为 PO'。只有 P_S 和 P_E 两点的位置不确定性才会引起线目标方向关系的不确定性，如图 5.20 中的 P_2 和 P_5。因此，若线目标 PO 误差带的宽度为 2ε，则线 PO 的粗集表达如图 5.20(b)所示，公式表达为：

$$\begin{cases} \overline{PO} = PO' \cdot \varepsilon \\ \mathrm{bn}(PO) = (P_S \cdot \varepsilon) \cup (P_E \cdot \varepsilon) \\ \underline{PO} = \overline{PO} - \mathrm{bn}(PO) \end{cases} \tag{5.14}$$

当源目标 PO 为点时，若其点位误差为 ε ，则 PO 的粗集表达如图 5.21(a)所示，公式表达为：

$$\begin{cases} \underline{PO} = \varnothing \\ \overline{PO} = PO \cdot \varepsilon \\ \mathrm{bn}(PO) = \overline{PO} \end{cases} \tag{5.15}$$

当源目标 PO 为面时，若其面位误差环的宽度为 2ε ，则面 PO 的粗集表达如图 5.21(b)所示，公式表达为：

$$\begin{cases} \underline{PO} = PO \Theta \varepsilon \\ \overline{PO} = PO \cdot \varepsilon \\ \mathrm{bn}(PO) = \overline{PO} - \underline{PO} \end{cases} \tag{5.16}$$

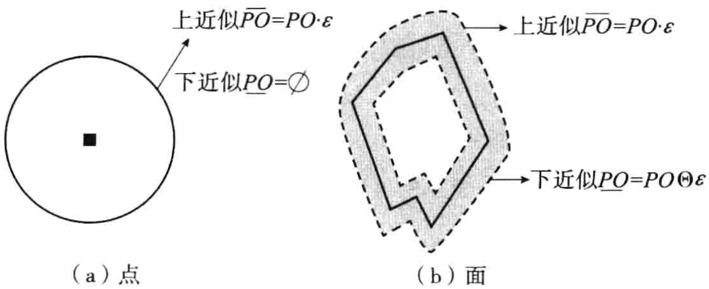


图 5.21 锥形模型中点、面目标位置不确定性的粗集表达

公式(5.14)、(5.15)、(5.16)中，• 表示目标形态膨胀算子，Θ 表示目标形态侵蚀算子，$\overline{PO}$ 均表示源目标粗集表达的上近似、$\underline{PO}$ 均表示源目标粗集表达的下近似，bn(PO) 均表示源目标粗集表达的边界。

公式(5.16)定义的线目标的粗集表达对于进一步推导线目标

的不确定方向关系是有意义的，但不能用于描述线目标的位置不确定性。

(2)锥形模型中空间方向关系的粗集表达

将方向关系锥形模型用粗集进行解释，锥形模型所分解的平面空间即为粗集中的论域空间，具体方向的划分方法即为粗集论域集合上的划分函数，锥形模型所划分的各主方向即为粗集知识(库)。方向关系锥形模型的粗集描述如定义 5.1。

定义 5.1：设论域空间 U 为以 O 为圆心的圆形区域(360°)，R 是 U 上的一族基于方向划分的等价关系，令 $K=(U, R)$ 为方向关系知识库，$U/R=\{O_1, O_2, \cdots, O_N\}$，$O_n(n=4$ 或 8 或 16 或 $\cdots)$ 为基于锥形模型以 O 为参考点对 U 划分的 n 个方向区域，当 $n=4$ 时，$U/R=\{$N, E, S, W$\}$，方向关系知识库中每个元素对应的圆心角为 90°；当 $n=8$ 时，$U/R=\{$N, NE, E, SE, S, SW, W, NW$\}$，方向关系知识库中每个元素对应的角度为 45°；当 $n=16$ 时，$U/R=\{$N, NNE, NE, ENE, E, ESE, SE, SSE, S, SSW, SW, WSW, W, WNW, NW, NNW$\}$，方向关系知识库中每个元素对应的角度为 22.5°……

在锥形模型中，如果源目标落在某一方向的锥形区域内，则认为源目标在参考目标的主方向上。因此，定义源目标 PO 隶属于锥形模型基于参考点 O 所划分的 n (n 为方向分辨率，$n=4$ 或 8 或 16 或 $\cdots$)个方向区域 $O_i(1 \leqslant i \leqslant n)$ 的隶属函数为：

$$u(PO \in O_i)=\begin{cases}1 & PO \cap O_i = t \\ 0 & PO \cap O_i = f\end{cases} \tag{5.17}$$

参考目标(点) O 和源目标 PO 之间的方向关系是：

$$R_{O(PO)}=\{O_i \mid u(PO \in O_i)=1\} \tag{5.18}$$

定义 5.2：源目标 PO 相对于参考目标(点) O 的方向关系 $R_{O(PO)}$ 的下粗近似集为 $\underline{R}_{O(PO)}$ 、上粗近似集为 $\overline{R}_{O(PO)}$，$\underline{R}_{O(PO)}=R_{O(\underline{PO})}=\{O_i \mid u(\underline{PO} \in O_i)=1\}$，$\overline{R}_{O(PO)}=R_{O(\overline{PO})}=\{O_i \mid u(\overline{PO} \in O_i)=1\}$，$R_{O(PO)}$ 的边界集合 $\mathrm{bn}(R_{O(PO)})=\overline{R}_{O(PO)}-\underline{R}_{O(PO)}$。

具体计算中，先确定方向分辨率 n，再结合定义 5.2 利用公式(5.17)、(5.18)确定空间目标不确定方向关系的粗集表达结果。

2. 方向关系矩阵模型中空间方向关系粗集表达

杜世宏等人(2003)将模糊面状目标和精确面状目标用统一的粗集表达模型进行表达，并在空间方向关系矩阵模型中构建了面状目标间空间方向关系的粗集表达方法。

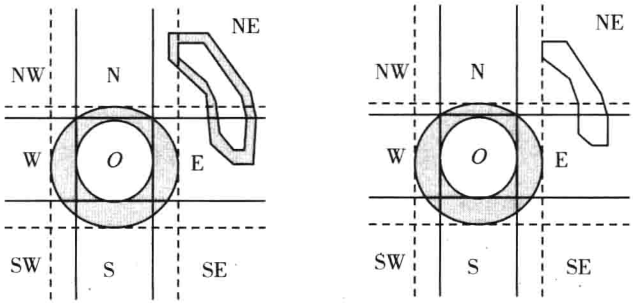
该方法首先定义模糊面状目标的粗集表达为：

$$\underline{A}_{\alpha} = \{(x, y) \in U \mid u(x, y) \geqslant \alpha\} \tag{5.19}$$

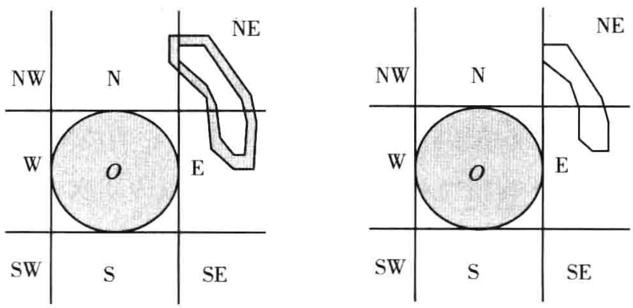
$$\overline{A}_{\beta} = \{(x, y) \in U \mid u(x, y) \geqslant \beta\} \tag{5.20}$$

式(5.19)和(5.20)中，U 为 2 维元素 (x, y) 的论域空间，A 为 2 维模糊集合(对象)，$u(x, y)$ 为其隶属函数，对于任意的实数 $0 < \beta \leqslant \alpha \leqslant 1$，$\underline{A}_{\alpha}$ 为模糊目标 A 的下粗近似集，$\overline{A}_{\beta}$ 为模糊目标 A 的上粗近似集。当 A 为精确对象时，$u(x, y)$ 取值为 0 或 1，令 $\beta = 0$，$\alpha = 1$，则 2 维模糊集 A 的下粗、上粗近似集分别与 A 相等，即 $\underline{A}_{\alpha} = \overline{A}_{\beta} = A$，$\mathrm{bn}R(A) = 0$，这与 A 是精确对象的概念是一致的。

然后定义模糊对象 A 和 B 之间的方向关系 O_{AB} 的下粗近似集、上粗近似集为$\underline{O}_{AB}$和$\overline{O}_{AB}$，$\underline{O}_{AB} = \{O_{\underline{AB}}\}$，$\overline{O}_{AB} = \{O_{\underline{A}\bar{B}}, O_{\bar{A}\underline{B}}, O_{\overline{AB}}\}$。如图 5.22 所示，对象 A 的$\underline{A}_{\alpha}$和$\overline{A}_{\beta}$ 的外接矩形分别把 A 的周围空间区域划分为 9 个区域$\underline{O}_{i}$和$\overline{O}_{i}(1 \leqslant i \leqslant 9)$，根据$\underline{O}_{i}$ 和$\overline{O}_{i}$ 与对象 B 的下、上粗近似集的关系可以确定 A 和 B 之间的方向关系有 4 种($O_{\underline{AB}}$，$O_{\underline{A}\bar{B}}$，$O_{\bar{A}\underline{B}}$，$O_{\overline{AB}}$)。$O_{\underline{AB}}$ 表示对象 A 的下粗近似集$\underline{A}_{\alpha}$ 与对象 B 的下粗近似集$\underline{B}_{\alpha}$之间的方向关系，$O_{\underline{A}\bar{B}}$ 表示对象 A 的下粗近似集$\underline{A}_{\alpha}$ 与对象 B 的上粗近似集$\overline{B}_{\beta}$ 之间的方向关系，其他类推。由于$O_{\underline{AB}}$ 的确定性和精确性程度大于其他的关系，因而对象 A，B 之间的方向关系的下粗近似集 $\underline{O}_{AB}$ 由 $O_{\underline{AB}}$ 组成，而其上粗近似集 $\overline{O}_{AB}$ 由 $O_{\underline{A}\bar{B}}$，$O_{\bar{A}\underline{B}}$，$O_{\overline{AB}}$ 的并集组成，即$\overline{O}_{AB} = \{O_{\underline{A}\bar{B}}, O_{\bar{A}\underline{B}}, O_{\overline{AB}}\}$。



(a) 模糊对象与模糊对象间方向关系　(b) 模糊对象与精确对象间方向关系



(c) 精确对象与模糊对象间方向关系　(d) 精确对象与精确对象间方向关系

图 5.22　面状对象间方向关系粗集表达

3. 多粒度空间方向关系不确定性的定量评价

关于空间方向关系模型，方向系统的划分存在多粒度性，如锥形模型中，有 4 方向、8 方向、16 方向等，而且方向系统划分粒度越细，方向关系越准确。

粗集的不确定性主要由两个原因引起：一个原因来自于给定近似空间的粗集的边界，这种不确定性称为系统的不确定性，用近似精度和粗糙度进行度量；另一个原因直接来自于论域上二元关系对论域中对象进行分类的能力，即知识的粒度，知识粒度越大，相对于近似空间中的概念和知识就越不确定，这种不确定性称为概念的不确定性，用基于知识含量的近似精度的粗糙度进行度量。与之对

应，在空间方向关系的粗集表达中，空间数据位置不确定性引起的方向关系粗集表达的不确定性相当于粗集的系统不确定性，方向系统划分粒度引起的空间方向关系的不确定性相当于粗集的概念不确定性。因此，基于空间方向关系不确定性的粗集表达定义了近似精度与粗糙度用于度量空间数据位置不确定性引起的方向关系粗集表达的不确定性；构建了知识含量的测度用于度量锥形模型粗集表达的方向分类能力；并在近似精度和粗糙度中引入知识含量测度因子，构建了基于知识含量的近似精度和粗糙度，用于确定性地理目标位置不确定性及方向系统划分粒度引起的方向关系不确定性的定量评价。

(1)近似精度与粗糙度

定义 5.3：

$$\alpha_R(R_{O(PO)}) = \frac{|\underline{R}_{O(PO)}|}{|\overline{R}_{O(PO)}|} \tag{5.21}$$

$$\rho_R(R_{O(PO)}) = 1 - \alpha_R(R_{O(PO)}) \tag{5.22}$$

分别被称为空间对象 PO 相对于参考点 O 的方向关系集合 $R_{O(PO)}$ 的近似精度和粗糙度。

公式(5.21)中，$|\underline{R}_{O(PO)}|$ 表示方向关系上粗近似集 $\overline{R}_{O(PO)}$ 中元素的个数，$|\overline{R}_{O(PO)}|$ 表示方向关系下粗近似集 $\underline{R}_{O(PO)}$ 中元素的个数。

$\alpha_R(R_{O(PO)})$ 和 $\rho_R(R_{O(PO)})$ 都是一个区间 $[0,1]$ 上的实数，$\alpha_R(R_{O(PO)})$ 越大，集合 $R_{O(PO)}$ 的确定度就越大；$\rho_R(R_{O(PO)})$ 越大，集合 $R_{O(PO)}$ 的不确定度就越大。

$\alpha_R(R_{O(PO)})$ 和 $\rho_R(R_{O(PO)})$ 用来度量空间数据位置不确定性引起的方向关系粗集表达的不确定性。

(2) 知识含量的度量

方向关系的粗集表达中，方向系统划分的粒度即为粗集中知识的粒度。知识 R 含量的测度用于度量锥形模型粗集表达的方向分类能力。本文中的知识均指方向系统划分产生的主方向。

定义 5.4：对于基于锥形模型的方向关系知识库 $K = (U, R)$，

$U/R=\{O_1,O_2,\cdots,O_n\}$，称

$$I(R_n)=1-\frac{1}{|U|^2}\sum_{i=1}^{n}|O_i|^2 \tag{5.23}$$

为知识 R 含量的测度。

式(5.23)中，$|U|=360$，在 4 方向系统中，即 $n=4$ 时，$|O_i|=90$，$I(R_4)=0.75$；在 8 方向系统中，即 $n=8$ 时，$|O_i|=45$，$I(R_8)=0.875$；在 16 方向系统中，即 $n=16$ 时，$|O_i|=22.5$，$I(R_{16})=0.938$；在 32 方向系统中，即 $n=32$ 时，$|O_i|=11.25$，$I(R_{32})=0.969$；在 64 方向系统中，即 $n=64$ 时，$|O_i|=5.625$，$I(R_{64})=0.984$……；用方位角描述方向关系，即 $n=360$ 时，$|O_i|=1$，$I(R_{360})=0.997$。方向系统划分粒度 n 与知识 R 含量的测度 $I(R_n)$ 的对应关系见表 5.13。

表 5.13　**方向系统划分粒度 n 与知识 R 含量的测度 $I(R_n)$ 的对应关系**

n	4	8	16	32	64	…	360
$\lvert O_i \rvert$	90	45	22.5	11.25	5.625	…	1
$I(R_n)$	0.75	0.875	0.938	0.969	0.984	…	0.997

通过表 5.13 可以看出，$I(R_n)$ 随方向系统划分粒度 n 的增大而增大，这与“n 越大，方向划分越准确”是一致的。

(3)基于知识含量的近似精度与粗糙度

基于知识含量的近似精度和粗糙度是在近似精度和粗糙度中引入知识含量的测度因子 $I(R_n)$，克服了近似精度和粗糙度不能全面衡量粗集不确定性的缺点。

定义 5.5：空间对象 PO 相对于参考点 O 的方向关系为 $R_{O(PO)}$，称

$$\mathrm{Rough}_R(R_{O(PO)})=\left(1-\frac{|\underline{R_{O(PO)}}|}{|\overline{R_{O(PO)}}|}\right)(1-I(R_n)) \tag{5.24}$$

$$D_R(R_{O(PO)})=I(R_n)+\alpha_R(R_{O(PO)})-I(R_n)\alpha_R(R_{O(PO)}) \tag{5.25}$$

分别为 $R_{O(PO)}$ 关于 U 上的等价关系 R 的基于知识(方向系统划分粒度)R 含量的粗糙度和近似精度。

式(5.24)中,$|\underline{R}_{O(PO)}|$ 表示方向关系上粗近似集 $\overline{R}_{O(PO)}$ 中元素的个数,$|\overline{R}_{O(PO)}|$ 表示方向关系下粗近似集 $\underline{R}_{O(PO)}$ 中元素的个数。

$\mathrm{Rough}_R(R_{O(PO)})$ 和 $D_R(R_{O(PO)})$ 都是一个区间[0,1]上的实数,空间方向关系系统划分粒度越小,$I(R)$ 越大,$D_R(R_{O(PO)})$ 越大,$\mathrm{Rough}_R(R_{O(PO)})$ 越小,集合 $R_{O(PO)}$ 的确定度就越大,不确定度就越小,反之亦然。$\mathrm{Rough}_R(R_{O(PO)})$ 和 $D_R(R_{O(PO)})$ 用于综合评价由于粗集边界(位置不确定性引起)和方向系统划分粒度引起的方向关系的不确定性。

4. 实例与讨论

按照本节的研究,多粒度空间方向关系不确定性的定量评价流程如图 5.23 所示。

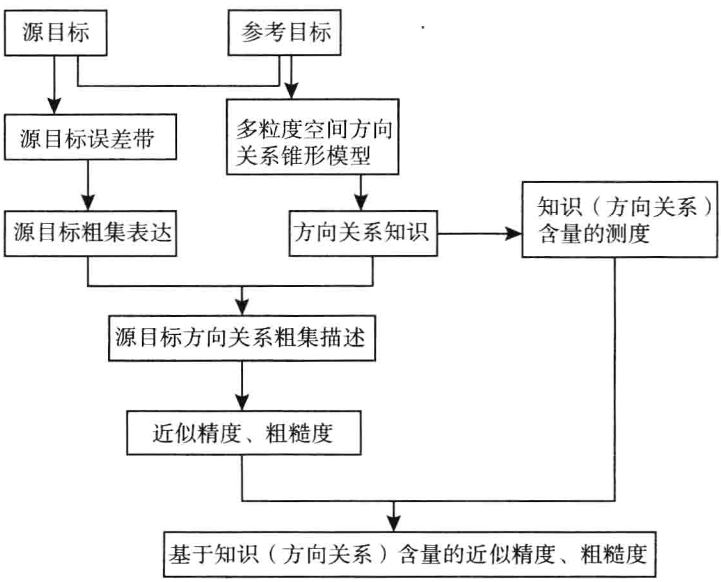


图 5.23 多粒度空间方向关系定量评价流程图

实验数据如图 5. 24 所示，参考点位置为参考目标的质心，源目标为目标 1（点目标）、目标 2（线目标）、目标 3（面目标）。依据本文的研究方法，首先对图 5. 24 中目标 1、目标 2、目标 3 按其精度指标计算其粗集表达结果，见表 5. 14，然后计算方向关系的不确定性评价的指标：近似精度、粗糙度与基于知识含量的近似精度、基于知识含量的粗糙度，见表 5. 15。

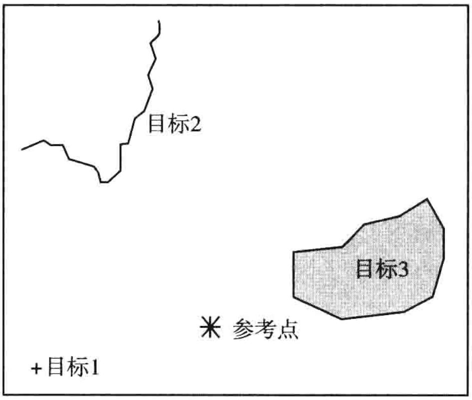


图 5. 24　实验数据

表 5. 14　　空间方向关系粗集表达结果

		上近似	下近似	边界集合
源目标 1	方位角	[253°,264°]	∅	[253°,264°]
	…	…	…	…
	16 方向系统	{W,WSW}	∅	{W,WSW}
	8 方向系统	{W}	∅	{W}
	4 方向系统	{W}	∅	{W}

续表

		上近似	下近似	边界集合
源目标 2	方位角	[305°,353°]	[312°,345°]	[305°,312°)∪(345°,353°]
	…	…	…	…
	16 方向系统	{NW,NNW,N}	{NW,NNW}	{N}
	8 方向系统	{NW,N}	{NW,N}	∅
	4 方向系统	{W,N}	{W,N}	∅
源目标 3	方位角	[47°,90°]	[64°,78°]	[47°,64°)∪(78°,90°]
	…	…	…	…
	16 方向系统	{NE,ENE,E}	{ENE}	{NE,E}
	8 方向系统	{NE,E}	{NE,E}	∅
	4 方向系统	{E}	{E}	∅

表 5.15 **源目标多粒度空间方向关系不确定性度量指标值**

		近似精度	粗糙度	基于知识含量的近似精度	基于知识含量的粗糙度
源目标 1	方位角	0	1	0.997	0.003
	…	…	…	…	…
	64 方向系统	0	1	0.984	0.016
	32 方向系统	0	1	0.969	0.031
	16 方向系统	0	1	0.938	0.062
	8 方向系统	0	1	0.875	0.125
	4 方向系统	0	1	0.75	0.25

续表

		近似精度	粗糙度	基于知识含量的近似精度	基于知识含量的粗糙度
源目标 2	方位角	0. 688	0. 312	0. 999	0. 001
	…	…	…	…	…
	64 方向系统	0. 7	0. 3	0. 995	0. 005
	32 方向系统	0. 8	0. 2	0. 994	0. 006
	16 方向系统	0. 667	0. 333	0. 98	0. 02
	8 方向系统	1	0	1	0
	4 方向系统	1	0	1	0
源目标 3	方位角	0. 326	0. 674	0. 998	0. 002
	…	…	…	…	…
	64 方向系统	0. 444	0. 556	0. 991	0. 009
	32 方向系统	0. 4	0. 6	0. 981	0. 019
	16 方向系统	0. 333	0. 667	0. 96	0. 04
	8 方向系统	1	0	1	0
	4 方向系统	1	0	1	0

通过对表 5. 15 中数据的分析，随着方向系统划分粒度的增大，近似精度和粗糙度的变化没有规律，这与人类空间认知逻辑不一致，这充分说明了近似精度和粗糙度只能部分反映粗集的不确定性。而基于知识含量的近似精度随方向系统划分粒度的增大而减小，基于知识含量的粗糙度随方向系统划分粒度的增大而增大，说明基于知识含量的近似精度和粗糙度对方向系统划分粒度的变化是敏感的，是评价由于目标位置不确定性及方向系统划分粒度引起的方向关系不确定性的理想指标。

因为实验中线目标和面目标不确定边界方向分布范围比 4 方向系统和 8 方向系统主方向范围小很多，使得线及面目标方向关系在 4 方向及 8 方向系统中的粗集表达的上近似和下近似相等，从而导

致表5.15中标有下划线的数字不符合变化规律。

5.3.3 多尺度空间方向关系不确定性建模

空间数据尺度转换的结果必然使空间数据的不确定性发生变化，不同尺度的空间数据其不确定性不同，致使其对应的空间方向关系也有所差异，因此，有必要研究多尺度空间方向关系的不确定性问题。对应多尺度空间方向关系的不确定性问题，一方面建立不同尺度下空间方向关系的不确定性描述，如蔡剑红(2007)建立的多尺度下顾及不确定性的空间方向关系进行形式化描述模型。另一方面，考虑由于多尺度空间目标位置不确定性引起多尺度空间方向关系之间的相似或差异程度，本研究构建了基于位置不确定性的多尺度空间方向关系相似性的粗集模型。

1. 多尺度面目标间空间方向关系的粗集表达

当用锥形模型来描述面目标间的方向关系时，首先计算参考目标的质心作为锥形模型的中心(参考点)，假定参考目标是均质的，参考目标节点坐标为 $(x_1, y_1, x_2, y_2, \cdots, x_{s-1}, y_{s-1}, x_s, y_s)$，$s$ 为参考目标节点的个数，$s \geqslant 1$，其质心 $O(x_0, y_0)$ 的计算公式为：

$$\begin{cases} x_0 = \sum_{j=1}^{s-1} (I_{j,j+1} \times x_{j,j+1}) \Big/ \sum_{j=1}^{s-1} I_{j,j+1} \\ y_0 = \sum_{j=1}^{s-1} (I_{j,j+1} \times y_{j,j+1}) \Big/ \sum_{j=1}^{s-1} I_{j,j+1} \end{cases}$$

上式中，$x_{j,j+1} = (x_j + x_{j+1})/2$，$y_{j,j+1} = (y_j + y_{j+1})/2$。$I_{j,j+1} = \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2}$。

以 $O(x_0, y_0)$ 为空间参考点，将周围的区域按多种方向系统分别划分成若干个具有方向性的子区域。

依据方向关系锥形模型的粗集描述定义(定义5.1)，进一步定义源目标(平面多边形) A 隶属于锥形模型基于参考点 O 所划分的 n (n 为方向分辨率，$n = 4$ 或 8 或 16 或 $\cdots$)个方向区域 $O_i (1 \leqslant i \leqslant n)$ 的隶属函数为：

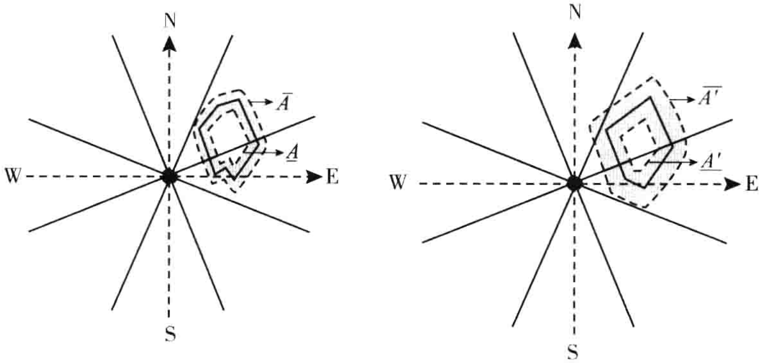
$$\mu_i = \frac{\text{Area}(A \cap O_i)}{\text{Area}(A)} \tag{5.26}$$

对象 A 相对于 O 的方向关系为：

$$R_{OA}=\{\mu_i \mid \mu_i=\frac{\text{Area}(A\cap O_i)}{\text{Area}(A)}\} \tag{5.27}$$

具有位置不确定性的对象 A 相对于参考点 O 的方向关系 R_{OA} 的下粗近似集 $\underline{R}_{OA}=R_{O\underline{A}}$、上粗近似集 $\overline{R}_{OA}=R_{O\overline{A}}$、边界集合 $\text{bn}(R_{OA})=\overline{R}_{OA}-\underline{R}_{OA}$。如图 5.25(a)所示，$\underline{R}_{OA}=\{\mu_i \mid \mu_i=\frac{\text{Area}(\underline{A}\cap O_i)}{\text{Area}(\underline{A})}\}$，

$\overline{R}_{OA}=\{\mu_i \mid \mu_i=\frac{\text{Area}(\overline{A}\cap O_i)}{\text{Area}(\overline{A})}\}$。



（a）尺度1　　（b）尺度2

图 5.25　多尺度面目标空间方向关系不确定性的粗集表达

按照此方式，可以得到任意尺度上空间方向关系不确定性的粗集表达。图 5.25（b）是图 5.25(a)简化的结果，图 5.25(b)所示平面 A' 相对于参考点 O 的方向关系 $R_{OA'}$ 的下粗近似集 $\underline{R}_{OA'}=R_{O\underline{A'}}=\{\mu_i \mid \mu_i=\frac{\text{Area}(\underline{A'}\cap O_i)}{\text{Area}(\underline{A'})}\}$、上粗近似集 $\overline{R}_{OA'}=R_{O\overline{A'}}=\{\mu_i \mid \mu_i=\frac{\text{Area}(\overline{A'}\cap O_i)}{\text{Area}(\overline{A'})}\}$、边界集合 $\text{bn}(R_{OA'})=\overline{R}_{OA}-\underline{R}_{OA'}$。

具体计算中，先确定方向分辨率 n，再利用式(5.26)、式

(5.27)确定空间目标方向关系不确定性的粗集表达结果。

2. 基于位置不确定性的多尺度空间方向关系相似性粗集模型

(1)空间方向相似性的定义与计算

Goyal(2000)基于方向关系矩阵模型定义了面目标间方向关系的相似性及其计算方法。方向关系矩阵模型通过记录源目标对象落在基于参考对象划分的哪一个方向片来描述方向。如图5.26所示，将图5.26(a)作为参考区域，再把图5.26(b)、图5.26(c)和图5.26(d)分别与图5.26(a)比较，并计算其方向相似性。

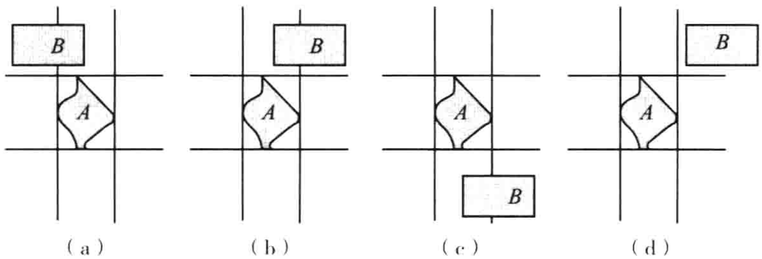


图5.26 面状对象 A 和 B 间的方向关系(A 为参考对象)

该方法最关键有四步，一是定义空间方向距离(distance between two cardinal direction)，并建立方向片的临近方向距离图；二是计算源目标位于各方向片面积比的方向关系矩阵；三是确定两个方向关系矩阵最小转换代价；四是定义并计算空间方向关系相似性值。

Tversky 早在1977年进行相似性研究时就指出相似性是一种直觉和主观判断，它没有严格的数学模型(Roop K. Goyal，2000)。那么相似值就不可能有一个精确的值，我们只能从相似值中来比较相似程度。因为空间方向相似性在认知上和应用上就有一定的模糊性，所以可以在一定精度范围内来计算空间方向相似性值。同时，Goyal(2000)提出的用最优化方法计算两个多要素方向关系矩阵的最小转换代价比较复杂，郭庆胜等人(2004)采用栅格数据来表达面状目标，把多要素方向关系矩阵转换为单要素方向关系矩阵，使

空间方向相似性值的计算简单化。面状目标用栅格数据描述后，就可以明确定义每个栅格单元位于哪一个方向片，一个面状目标相对于参考面状目标的方向关系矩阵就可以转换为该面状目标的所有栅格单元的方向关系矩阵的集合。

这里在基于方向关系矩阵模型定义的方向关系相似性的基础上，定义了基于锥形模型的空间目标相似性的计算方法。基于锥形模型的多尺度空间方向关系相似性的定义与计算过程与基于方向关系矩阵模型的方向关系相似性的定义与计算过程一致，具体分为以下四步：

①临近方向距离的定义。方向关系锥形模型对方向片的划分有 4 方向系统、8 方向系统、16 方向系统等，4、8、16 方向系统的临近方向距离图如图 5.27 所示。基于方向关系矩阵模型的方向关系相似性的计算过程中临近方向距离的计算借用了图像处理中的 4 连通原理，本书不考虑参考目标所占据的中心区域，从具体方向开始取顺时针和逆时针方向移动距离的最小值作为临近方向距离。

②源目标位于各方向片面积比的方向关系矩阵。空间方向关系的粗集表达其实质是源目标的上近似和下近似分别位于各方向片面积比的方向关系矩阵，在具体尺度上用 2 个 $1\times n$（$n=4$ 或 8 或 16）方向关系矩阵描述源目标相对于参考点的方向关系。例如，将图 5.25 中图形数据用定义 5.3 和式(5.27)进行运算，在 8 方向系统中，尺度 1 上的源目标位于各方向片面积比的两个方向关系矩阵为：$\underline{R}_{OA}=\{0, 0.8, 0.2, 0, 0, 0, 0, 0\}$、$\overline{R}_{OA}=\{0, 0.71, 0.29, 0, 0, 0, 0, 0\}$，尺度 2 上源目标位于各方向片面积比的两个方向关系矩阵为：$\underline{R}_{OA'}=\{0, 0.79, 0.21, 0, 0, 0, 0, 0\}$、$\overline{R}_{OA'}=\{0, 0.02, 0.87, 0.11, 0, 0, 0, 0\}$。

③两个方向关系矩阵最小转换代价的确定。与基于方向关系矩阵模型的方向关系相似性中两个方向关系矩阵最小转换代价的计算方法一样用最优化方法计算，先计算两个方向关系矩阵的差异矩阵，再根据差异矩阵计算两个方向关系矩阵的最小转换代价。本书中考虑的是多尺度表达中方向关系的相似性，因此，所要比较的方

	N	E	S	W
N	0	1	2	1
E	1	0	1	2
S	2	1	0	1
W	1	2	1	0

（a）4方向系统临近方向距离图

	N	NE	E	SE	S	SW	W	NW
N	0	1	2	3	4	3	2	1
NE	1	0	1	2	3	4	3	2
E	2	1	0	1	2	3	4	3
SE	3	2	1	0	1	2	3	4
S	4	3	2	1	0	1	2	3
SW	3	4	3	2	1	0	1	2
W	2	3	4	3	2	1	0	1
NW	1	2	3	4	3	2	1	0

（b）8方向系统临近方向距离图

	N	NNE	NE	ENE	E	ES	ESE	SSE	S	SSW	SW	WSW	W	WN	WNW	NNW
N	0	1	2	3	4	5	6	7	8	7	6	5	4	3	2	1
NNE	1	0	1	2	3	4	5	6	7	8	7	6	5	4	3	2
NE	2	1	2	1	2	3	4	5	6	7	8	7	6	5	4	3
ENE	3	2	1	0	1	2	3	4	5	6	7	8	7	6	5	4
E	4	3	2	1	0	1	2	3	4	5	6	7	8	7	6	5
ESE	5	4	3	2	1	0	1	2	3	4	5	6	7	8	7	6
SE	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	7
SSE	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8
S	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7
SSW	7	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6
SW	6	7	8	7	6	5	4	3	2	1	0	1	2	3	4	5
WSW	5	6	7	8	7	6	5	4	3	2	1	0	1	2	3	4
W	4	5	6	7	8	7	6	5	4	3	2	1	0	1	2	3
WNW	3	4	5	6	7	8	7	6	5	4	3	2	1	0	1	2
NW	2	3	4	5	6	7	8	7	6	5	4	3	2	1	0	1
NNW	1	2	3	4	5	6	7	8	7	6	5	4	3	2	1	0

（c）16方向系统临近方向距离图

图5.27 锥形模型中临近方向距离图

向关系矩阵是尺度1上源目标方向关系的上近似、下近似，分别与尺度2上源目标方向关系的上近似、下近似的比较，共有四种组

合。如图 5.25 在 8 方向系统中尺度 1 上 A 的方向关系矩阵到尺度 2 上 A' 的方向关系矩阵的差异矩阵分别为 $\Delta(\overline{R}_{OA},\overline{R}_{OA'})=\{0.02,0.16,-0.18,0,0,0,0,0\}$、$\Delta(\overline{R}_{OA},\underline{R}_{OA'})=\{0,0.08,-0.08,0,0,0,0,0\}$、$\Delta(\underline{R}_{OA},\overline{R}_{OA'})=\{0.02,0.07,-0.09,0,0,0,0,0\}$、$\Delta(\underline{R}_{OA},\underline{R}_{OA'})=\{0,-0.01,0.01,0,0,0,0,0\}$，于是在 8 方向系统中从 A 到 A' 的最小转换代价分别为 $\text{mincost}(\overline{R}_{OA},\overline{R}_{OA'})=0.02\times 2+0.16\times 1=0.2$，$\text{mincost}(\overline{R}_{OA},\underline{R}_{OA'})=0.08$、$\text{mincost}(\underline{R}_{OA},\underline{R}_{OA'})=0.01$、$\text{mincost}(\underline{R}_{OA},\overline{R}_{OA'})=0.11$。

④多尺度空间方向关系相似值的定义。初始尺度上 A 和 O 方向关系 R_{OA} 与简化尺度上 A' 和 O 的方向关系 $R_{OA'}$ 的相似性为 $S(R_{OA},R_{OA'})$，则

$$
S(R_{OA},R_{OA'})=\begin{cases}1-\dfrac{\text{mincost}(R_{OA},R_{OA'})}{2},\text{当方向系统分粒度为 4 时;}\\[2ex]1-\dfrac{\text{mincost}(R_{OA},R_{OA'})}{4},\text{当方向系统分粒度为 8 时;}\\[2ex]1-\dfrac{\text{mincost}(R_{OA},R_{OA'})}{8},\text{当方向系统分粒度为 16 时;}\end{cases}\tag{5.28}
$$

因为本书重点考虑多尺度空间方向关系的相似性，同时顾及了源目标的不确定性，尺度变化时，在具体划分粒度的方向系统中，$S(R_{OA},R_{OA'})$ 的取值有 4 种可能：$S(\underline{R}_{OA},\underline{R}_{OA'})$、$S(\underline{R}_{OA},,\overline{R}_{OA'})$、$S(\overline{R}_{OA},\underline{R}_{OA'})$、$S(\overline{R}_{OA},\overline{R}_{OA'})$。图 5.25(a)中目标 A 的上近似 $\overline{A}$ 与图 5.25(b)目标 A' 的上近似 $\overline{A'}$ 的在 8 方向锥形模型中方向关系相似性值为 $S(\overline{R}_{OA},\overline{R}_{OA'})=1-\dfrac{\text{mincost}(\overline{R}_{OA},\overline{R}_{OA'})}{4}=1-\dfrac{0.2}{4}=0.95$，同理，$S(\underline{R}_{OA},\underline{R}_{OA'})=0.998$、$S(\underline{R}_{OA},\overline{R}_{OA'})=0.973$、$S(\overline{R}_{OA},\underline{R}_{OA'})=0.98$。

(2)基于位置不确定性的多尺度空间方向关系相似性的粗集表达

尺度变化时，源目标的位置不确定性发生了变化，考虑源目标在不同尺度上的位置不确定性，在每个方向系统划分粒度下，$S(R_{OA},R_{OA'})$ 的取值有 4 种可能：$S(\underline{R}_{OA},\underline{R}_{OA'})$、$S(\underline{R}_{OA},\overline{R}_{OA'})$、$S(\overline{R}_{OA},\underline{R}_{OA'})$、$S(\overline{R}_{OA},\overline{R}_{OA'})$。

定义 5.6：在某一粒度划分的方向关系锥形模型中，$S(R_{OA},R_{OA'})$ 的 4 个取值大小可能不等，定义 4 个取值的最小值为多尺度方向关系相似性的上近似 $\overline{S}(R_{OA},R_{OA'})$、最大值为多尺度方向关系相似性的下近似 $\underline{S}(R_{OA},R_{OA'})$，多尺度方向关系相似性的边界集合 $\mathrm{bn}(S)$ 为由上近似和下近似所确定的开区间，也就是说

$$\underline{S}(R_{OA},R_{OA'})=\max\left(S(\underline{R}_{OA},\underline{R}_{OA'}),S(\underline{R}_{OA},\overline{R}_{OA'}),S(\overline{R}_{OA},\underline{R}_{OA'}),S(\overline{R}_{OA},\overline{R}_{OA'})\right) \tag{5.29}$$

$$\overline{S}(R_{OA},R_{OA'})=\min\left(S(\underline{R}_{OA},\underline{R}_{OA'}),S(\underline{R}_{OA},\overline{R}_{OA'}),S(\overline{R}_{OA},\underline{R}_{OA'}),S(\overline{R}_{OA},\overline{R}_{OA'})\right) \tag{5.30}$$

$$\mathrm{bn}(S)=(\overline{S}(R_{OA},R_{OA'}),\underline{S}(R_{OA},R_{OA'})) \tag{5.31}$$

于是对图 5.25，在 8 方向锥形模型中，

$$\underline{S}(R_{OA},R_{OA'})=\max(0.998,0.973,0.98,0.95)=0.998,$$

$$\overline{S}(R_{OA},R_{OA'})=\min(0.998,0.973,0.98,0.95)=0.95,$$

$$\mathrm{bn}(S)=(0.95,0.998)。$$

同理，可以计算图 5.25 在 4 方向和 16 方向关系锥形模型中方向关系相似性的上近似、下近似和边界集合。

因为空间目标的方向关系在其粗集表达的边界区域内是不确定的，所以以不同尺度上空间目标方向关系上近似和下近似的两两组合定义的多尺度空间方向关系相似性的粗集表达，其实质是定义了一个多尺度空间方向关系相似性的取值区间。

3. 实例与讨论

在多尺度空间目标方向关系相似性的自动评价中，可用 $(\overline{S}(R_{OA},R_{OA'}),\underline{S}(R_{OA},R_{OA'}))$ 定义的取值区间来定义评价标准；而且，空间方向关系相似性值在实际应用中有时并不需要很精确，可用

$(\overline{S}(R_{OA},R_{OA'}),\underline{S}(R_{OA},R_{OA'}))$ 定义的指标区间作为多尺度空间目标相似性值的范围。

5.3.3 节第(2)、(3)两部分分别对多尺度空间方向关系相似性粗集表达所涉及的各定义和计算方法进行了研究，经过梳理，多尺度空间方向关系相似性粗集表达的步骤为：

①分别计算源目标 A 和简化后源目标 A' 的不确定带。假设源目标 A 各节点精度（中误差）相同；简化后源目标 A' 各综合不确定性用本书第 4 章所建立的方法求取。用源目标 A 的中误差建立起误差带，用简化后源目标 A' 的综合不确定性建立不确定带。

②依据本节的定义计算源目标 A 和简化后源目标 A' 的粗集表达。

③按照定义 5.3 计算源目标 A 和简化后源目标 A' 在某一粒度锥形模型中的粗集表达，再按公式(5.27)计算 A 和 A' 粗集表达的比例表达形式。

④依据临近方向距离图，如图 5.27 所示，按照式(5.28)计算 A 和 A' 在具体方向系统中的 4 个方向关系相似性值。

⑤按照式(5.29)、(5.30)、(5.31)计算 A 和 A' 在具体方向系统中方向关系相似性值的下近似和上近似。

按照这个步骤，计算得到图 5.25(a)和图 5.25(b)方向关系的相似性粗集表达见表 5.16。

表 5.16　**多尺度空间方向关系相似性粗集表达示例**

	上近似	下近似	相似性取值区间
16 方向系统	0.946	0.968	(0.946,0.968)
8 方向系统	0.95	0.998	(0.95,0.998)
4 方向系统	0.86	1	(0.86,1)

根据表 5.16，图 5.25 中图形在 4 方向锥形模型中相似性取值区间的长度为 1−0.86=0.14、在 8 方向系统中相似性取值区间的长度为 0.998−0.95=0.048、在 16 方向系统中相似性取值区间的长度为

0.968−0.946=0.022。由此可见，在空间目标的多尺度表达中，方向系统划分的粒度越细，方向关系相似性的取值区间越短。这表明：在方向关系的多尺度表达中，方向系统划分的粒度越细，用本章的方法计算的空间目标的方向关系的相似性值越接近真实值。这与空间方向系统划分粒度越细，依据它所计算的空间目标的方向关系越精确是一致的。

5.4　多尺度空间关系不确定性的统一建模

5.4.1　多尺度空间关系的统一表达模型

对给定的空间参考目标 A、源目标 B，当确定了具体尺度(s) 上空间参考目标 A 和源目标 B 的拓扑关系 T_{AB}、空间方向关系 Dir_{AB} 和距离关系 $H(A,B)$，其空间关系 sR_{AB} 可以定义为：

$$sR_{AB} = (s \cup A \cup B \cup T_{AB} \cup \mathrm{Dir}_{AB} \cup H(A,B))$$

用一个有序 6 元组表示为：

$$sR_{AB} = (s, A, B, T_{AB}, \mathrm{Dir}_{AB}, H(A,B)) \tag{5.32}$$

式(5.32)即为多尺度表达中基本空间关系的统一表达模型。式(5.32)中 s 可定义为比例尺；A 为空间参考目标、B 为源目标，A、B 可分别为点、线或面实体，当比例尺 s 变化时，A、B 的类型可随之发生变化，用 D、L、A 分别表示点、线、面实体，源目标和参考目标可分别定义为：

$$A = \text{“}A_D\text{” or “}A_L\text{” or “}A_A\text{”}$$

$$B = \text{“}B_D\text{” or “}B_L\text{” or “}B_A\text{”}$$

拓扑关系 T_{AB}、方向关系 Dir_{AB} 和距离关系 $H(A,B)$ 需要依据具体的空间实体的位置数据进行计算，所选用的模型不同，得到的各空间关系的结论也有所差别，当比例尺 s 变化时，空间关系也会随之发生相应的变化。一般情况下应选用具有多尺度包容性的拓扑关系 T_{AB}、方向关系 Dir_{AB}和距离关系 $H(A,B)$ 的计算和描述模型。例如，拓扑关系的表达模型 RCC-5 模型、基于方向 Voronoi 图(DVDs)的方

向群模型和 Hausdorff 模型距离具备多尺度包容性。

5.4.2 多尺度空间关系不确定性的统一表达模型

5.4.1 节定义的多尺度空间关系的统一表达模型中,将拓扑关系、方向关系、距离关系统一在一个6元组中。这里在定义多尺度空间关系不确定性的统一表达模型时,不再单独考虑距离关系,而是在表达拓扑关系时,用拓扑与度量相结合关系代替拓扑关系。于是,式(5.32)改变为:

$$SR_{AB} = (S, A, B, T_{AB}, \mathrm{Dir}_{AB}) \tag{5.33}$$

式(5.33)中, T_{AB} 表达为:

$$T_{AB} = (\underline{T}(A,B), \overline{T}(A,B), D) \tag{5.34}$$

或

$$T_{AB} = (\underline{T}(A,B), \overline{T}(A,B), k) \tag{5.35}$$

式(5.34)和式(5.35)按本章5.2.4节的方法构建。

Dir_{AB} 表达为:

$$\mathrm{Dir}_{AB} = (\underline{\mathrm{Dir}}(A,B), \overline{\mathrm{Dir}}(A,B), \mathrm{Rough}_{\mathrm{Dir}}) \tag{5.36}$$

式(5.36)中 $\underline{\mathrm{Dir}}(A,B)$ 表示 A 与 B 方向关系的下近似, $\overline{\mathrm{Dir}}(A,B)$ 表示 A 与 B 方向关系的上近似, $\mathrm{Rough}_{\mathrm{Dir}}$ 表示方向关系基于知识含量的粗糙度,具体按本章5.3.2节的方法计算。

将式(5.34)、(5.36)代入式(5.33)得到:

$$SR_{AB} = (S, A, B, \underline{T}(A,B), \overline{T}(A,B), D), (\underline{\mathrm{Dir}}(A,B), \overline{\mathrm{Dir}}(A,B), \mathrm{Rough}_R)$$

5.5 本章小结

本章首先构建了多尺度空间面目标的粗集表达模型。其次,将空间拓扑关系与度量关系相结合,在概念邻域和拓扑距离的基础上,定义了扩展的拓扑距离,并利用粗集理论构建了多尺度空间拓扑关系不确定性的评价模型。然后,研究了多尺度空间方向关系不

确定性的评价方法，具体内容包括发展了空间方向关系不确定性的粗集表达模型，构建了基于知识含量的近似精度和粗糙度用于定量评价多粒度空间方向关系的不确定性，并构建了基于位置不确定性的多尺度空间方向关系相似性的粗集模型，用于定量评价多尺度空间方向关系的不确定性。最后，对多尺度空间拓扑关系和方向关系的不确定性进行了统一建模。

第6章　空间数据多尺度表达的不确定性可视化

眼睛是人获取信息的最主要的通道，研究表明，人通过视觉所获得的信息占获取信息总数的60%以上。在计算机技术飞速发展的今天，可视化已经成为科学计算与分析的核心内容之一。不确定性可视化是不确定性数据和分析结果的一种表现形式，已成为空间信息阅读、理解进而交互作用中最重要的工具和手段。不确定性可视化技术能够全面且从本质上把握住空间数据多尺度表达的不确定性基本特征，便于最迅速、形象地传递和接收它们，因此空间数据多尺度表达的不确定性离不开可视化。从视觉和空间认知的角度表现不确定性信息是不确定性建模的重要组成部分。本章将系统地研究空间数据多尺度表达不确定性可视化中的方法。

6.1　概　　述

空间数据及其多尺度表达中存在大量的不确定性，其中有些不确定可以根据误差建模领域内提出的不确定性度量方法进行评估（史文中，2005）。然而，有些不确定性则无法用此方法评估，特别是由人们的感知和心理活动所造成的不确定性。因此，研究不确定性的本质，实现对GIS数据及其表达的不确定性的度量与表示（如可视化）方法是目前必须解决的问题。Bertin（1983）、Goodchild（1994）、MacEachren（1994）、Fisher（1994）、史文中（2005）等人对GIS及其空间分析中的不确定性问题进行了研究，随着GIS的快速发展，空间数据多尺度表达已成为国内外学者研究的重点问题，如何从视觉和空间认知的角度表现不确定性信息是不确定性建模的重

要组成部分。

6.1.1 可视化

可视化是表达信息世界的重要方法和手段，它的发展与人类现代文明的启蒙以及测量、绘画和科技的发展一脉相承。在地图学、科学与工程制图中，可视化理念与技术已经应用和发展了数百年；科学计算可视化自1987年提出以来，在各工程和计算机领域得到了广泛的应用和发展。进入21世纪，现有的可视化技术已难以应对海量、高维、多源、动态数据的分析挑战，需要综合可视化、图形学、数据挖掘理论与方法，研究新的理论模型、新的可视化方法和新的用户交互手段，辅助用户从大尺度、复杂、矛盾甚至不完整的数据中快速挖掘有用的信息以便做出有效的决策(陈为等，2013)。

1. 可视化的概念

可视化就是将事物或现象变成可被视觉所感知的过程。

可视化(visualization)是利用计算机图形学、图像处理技术和人机交互技术，将采集、分析或模拟的数据映射成为容易识别和理解的图形、图像、视频或动画等形式，并允许用户对数据进行交互分析的理论、方法和技术。因此，可以将可视化看成传统的科学可视化和信息可视化的泛称，即处理的对象可以是任意数据来源、任意数据特性以及异构、异质数据的组合。

2. 可视化的意义

实现科学计算可视化将极大地提高科学计算的速度和质量，实现科学计算工具和环境的进一步现代化；由于它可将计算过程和结果用图形和图像直观、形象、整体地表达出来，从而使许多抽象的，难于理解的原理、规律和过程变得更容易理解，枯燥而冗繁的数据或过程变得生动有趣，更人性化；同时，通过交互手段改变计算的环境和所依据的条件，观察其影响，实现对计算过程的引导和控制。

6.1.2 视觉与认知理论

视觉是人类获取外部世界信息的最重要通道，人眼是一个高带

宽的并行视觉信号输入处理器，带宽高达每秒100兆字节(Ward，2010)。数据或分析结果的可视化旨在提供一种直观的可视界面，用户通过视觉感知器官获取经过可视编码的信息，经过大脑解码并形成认知，并在交互分析过程中洞悉信息的内涵，获取解决问题的方法。

1. 视觉感知与认知

(1)感知与认知的定义

感知是指客观事物通过人的感觉器官在人脑中形成的直接反映。人类感觉器官包括眼、鼻、耳以及遍布身体各处的神经末梢等，相应的感知能力分别成为视觉、嗅觉、听觉和触觉等。其中与可视化密切相关的主要是指视觉感知。与感知对应的概念是认知，认知心理学将认知过程看成是由信息的获取、分析、归纳、编码、概念形成、提取和使用等一系列阶段组成的按一定程序进行信息加工的系统。

心理学上的双重编码理论，即人类的感知系统由负责语言和其他非语言信息(特别是视觉感知方面)的两个子系统组成，此外，大脑对于视觉信息的记忆效果和记忆速度要好于对语言的记忆效果和速度。这也是可视化有助于数据信息表达的一个重要的理论基础。

感知心理学家通常将视觉分为低阶视觉和高阶视觉两种类型。低阶视觉与物体的物理性质有关，包括深度、形状、边界、表面材质等。高阶视觉包括物体的识别和分类，属于人类的认知能力的重要组成部分。其中，低阶视觉已经在信息可视化和可视分析的研究中得到了广泛的验证。

(2)格式塔理论

格式塔理论最基本的法则是简单精炼法则，认为人们在进行观察的时候，倾向于将视觉感知内容理解为常规的、简单的、相连的、对称的或有序的结构。同时，人们在获取视觉感知的时候，会倾向于将事物理解为一个整体，而不是将事物理解为组成该事物所有部分的集合。格式塔法则又称为完图法则，主要包括：贴近原则、相似原则、连续原则、闭合原则、共势原则、好图原则、对称

性原则、经验原则。因此格式塔的基本思想是：视觉形象首先是作为统一的整体被认知的，而后才以部分的形式被认知，也就是说，人们先看见一个构图的整体，然后才看见组成这一构图整体的各个部分。可视化设计必须遵循心理学关于感知和认知的理论研究成果。

(3)视觉感知的相对性

人类认知系统的工作原理决定于对所观察事物的相对判断。例如，在日常生活中，人们在描述一个物体的外观(如长度、高度、重量等)的时候通常会采用另外一个众所周知的物体为参照物。Weber定律描述了这一现象：人类感知系统将可察觉的刺激强度的变化表达为一个目标刺激强度的百分比。在可视化中，通过相对判断而精确地揭示数据尺度等信息需要一定的前提条件。一般而言，如果物体使用相同的参照物或者相互对齐，则会有助于人们做出更加准确的相对判断。

2. 视觉通道

可视化将数据属性采用不同的视觉通道进行编码。由于各个视觉通道的特性的差异，当可视化的结果呈现于用户时，用户获取信息的难度和所需要的时间不尽相同。人类感知系统在获取周围信息的时候，存在两种最基本的感知模式。第一种模式感知的信息是对象的本身特征和位置等，对应的视觉通道类型为定性或分类。第二种感知的信息是对象的某一属性的取值大小，对应的视觉通道类型为定量或定序。

通常认为空间、标记、位置、尺寸、颜色、亮度、饱和度、色调、配色方案、透明度、方向、形状、纹理和动画是常用的视觉通道类型。某些视觉通道被认为属于定性的视觉通道，如形状、颜色的色调或空间位置，而大部分的视觉通道更加适合于编码定量的信息，如直线长度、区域面积、空间体积、斜度、角度、颜色的饱和度和亮度等。当然，视觉通道的类型不具有明确的界限。

视觉通道的类型决定了可视化不同的数据可能采用的视觉通道，而视觉通道的表现力和有效性则指导可视化设计者如何挑选合适的视觉通道，实现对数据信息完整而具有目的性地展现。

6.1.3　空间数据多尺度表达不确定性的可视化

不确定性的可视化研究是进入20世纪90年代以后随着计算机技术的发展而兴起的。这方面的研究目前主要有空间矢量数据误差模型的可视化表示、影像分类不确定性的可视化表示、GIS应用系统中不确定性的可视化等。其中，具有代表性的有：L. Bastinado, Jwork和P. F. Fisher等人对多分辨率卫星影像进行专题模糊分类时的不确定性可视化；Keith C. Clarke等人在NCGIA对数字制图的不确定性采用虚拟现实模型进行可视化表达；Lgor Drecki等人在采用Spot卫星影像数据进行土地利用分类时对地理数据的不确定性进行可视化表示；Tinghua Ai和Zhongliang Cai等人采用粒子系统模型、蒙特卡洛模拟法等在制作专题图时对多边形填充的不确定性进行动画模拟；刘大杰、童小华、刘文宝等人对空间数据中点、线、面的不确定性模型进行了可视化表达；Jian huaWang等人在GIS的应用系统中加入了不确定性可视化图标和界面。

人们对空间地理现象或目标的认识，由于受到人眼和观察器具的影响，总是限制在一定的空间精度和空间外延上，使可视分析与尺度紧密联系在一起。

1. 概念

空间数据多尺度表达的不确定性可视化的学科范围是地理信息科学，研究对象是空间数据的多尺度表达，其可视化的内容是地理信息的获取、处理和分析过程中产生的不确定性。因此与之相适应的，其可视化的要求还有对研究对象的抽象和综合，有数据在不同尺度下通过符号表达的过程。因此，从本质上讲空间数据多尺度表达的不确定性可视化是科学可视化的重要组成部分，但是在应用范畴和可视化要求方面有所差别，在实现方法技术上也有一些显著的差别。因此，空间数据多尺度表达的不确定性可视化可以定义为：空间数据多尺度表达的不确定性可视化是综合利用地图学、计算机图形学、图像处理技术和人机交互技术，将空间数据多尺度表达过程中的不确定性通过将分析或模拟的数据映射为容易识别和理解的图形、图像、视频或动画等形式，并允许用户对数据进行交互分析

的理论、方法和技术。因此，可以将可视化看成传统的科学可视化和信息可视化的泛称，即处理的对象可以是任意数据来源、任意数据特性以及异构、异质数据的组合。

采用声音、触觉、嗅觉、味觉等多种媒体方式可以使空间信息的传递、接收更为形象、具体和逼真，但是暂时看来，有的对地理空间信息意义并不大，如嗅觉、味觉、触觉等媒体渠道，而声音、音频媒体方式也主要起辅助作用，因此有的学者把可听、可嗅、可味、可触也归入可视化(狭义的理解上应不属可视化范畴)，目前我们也把它列入可视化范畴。因此，空间数据多尺度表达的不确定性可视化是科学计算可视化在地理信息科学领域的特定发展。

2. 形式

通常将空间数据多尺度表达的不确定性用图形或文本进行标示，这也是不确定性的最为简单也最为常用的形式。随着计算机科学与技术的发展，可视化出现了更多的形式；20 世纪 90 年代随着多媒体的发展与普及，使空间信息可视化进入一个崭新的时期，可视化的形式也更为多样，呈现多维化的局面，并不断地发展着。目前，空间数据多尺度表达的不确定性可视化形式主要有：误差椭圆与误差带、灰度图、色彩图、地图符号法、等值线法、三维可视化、动态可视化和多媒体可视化。

虚拟现实是空间数据多尺度表达的不确定性可视化进一步研究和发展的新方式，它是由计算机及其相应的硬件设备组成的高级人机交互系统，以视觉为主，同时结合听觉、触觉、嗅觉甚至味觉来感知的环境，使人们有如进入真实的环境中并与之交互作用。虚拟现实也是未来不确定性可视化的方式之一。

6.2 空间数据多尺度表达的不确定性可视化设计与流程设计

6.2.1 空间数据多尺度表达的不确定性可视化设计

空间数据多尺度表达的不确定性可视化的方法众多，这为实现

不同风格的可视化提供了方便，但是也为寻求最佳可视化带来了一定的困难。一般在进行空间数据多尺度表达的不确定性可视化设计的时候要遵循以下原则：

①科学有效：在选择可视化的方式时，要利用合适的可视化方法或其组合，在短时间内把空间数据多尺度表达过程中的不确定性传递给用户。

②简洁易用：在有限的现实空间里，简洁的可视化能够表达更多的数据，易于理解。在可视化设计时还要针对用户的特点和需求，设计简洁易用的可视化形式，可以让用户能够尽快地掌握交互处理方式。

③美观：可视化的美感侧重点不是视觉美感，而是对数据内涵的揭示，在可视化中视觉美感包括标记设计上的美感，空间的合理应用，色彩的协调搭配等。

6.2.2　空间数据多尺度表达的不确定性可视化流程

空间数据多尺度表达的不确定性可视化不是一个单独的算法，而是一个流程。除了视觉映射外，也需要设计并实现其他关键环节如前端的数据采集、处理和后端的用户交互。这些环节是解决实际问题必不可少的步骤，并且直接影响最终的可视化效果。作为可视化应用的设计者，解析可视化流程有助于把问题化整为零，降低设计的复杂度。作为可视化开发者，解析可视化流程有助于软件开发模块化，提高开发效率，缩小问题范围、重复利用代码。作为可视化软件工具开发者，解析可视化流程有助于设计工具库、编程界面和软件模块。可视化流程以数据流向为主线，其主要模块包括数据采集、数据处理与转换、可视化映射和映射感知。整个可视化过程可以看成数据流经一系列处理模块并得到转换的过程。用户通过可视化交互和其他模块互动，通过反馈提高可视化的效果。下面给出了空间数据多尺度表达的不确定性可视化的概念图。

空间数据多尺度表达的不确定性可视化流程设计如图6.1所示。

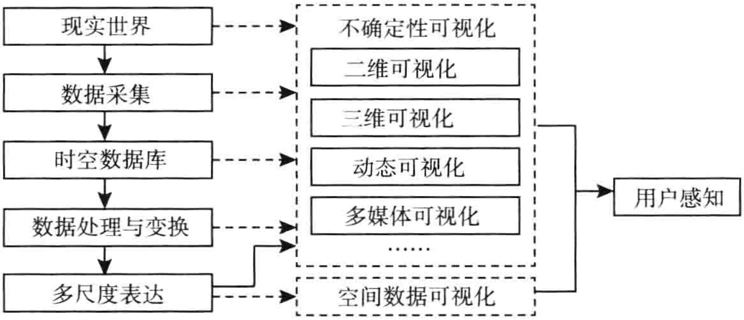


图 6.1 空间数据多尺度表达的不确定性可视化实现的流程

6.3 二维可视化

二维图形仍然是地理信息表达的重要手段，因此二维可视化也是空间数据不确定性表达的最基本的方法。目前，空间数据不确定性表达的主要方法有：误差椭圆与误差带、灰度图、色彩图、地图符号和等值线法等，本节将重点介绍空间数据多尺度表达不确定性的二维可视化。

6.3.1 误差椭圆与误差带

误差椭圆和误差带是最为常用的不确定性可视化模型。

1. 误差椭圆

在航空摄影测量、大地测量以及工程测量的误差理论中，椭圆常用于描述二维空间中点元的位置不确定性，该椭圆被称为误差椭圆。误差椭圆方法是基于位置不确定性服从正态分布的假设条件下推导出来的，对于 GIS 中的空间数据，基于中心极限定理，该假设可以认为通常是成立的（刘大杰，1999；史文中，2005）。空间数据多尺度表达中的原始空间数据可能是通过多次观测获得的，且每一次观测值都含有量值相约的随机误差，在制图综合或多尺度表达过程中又增加了误差，并在一定程度上进行累积，一般认为在空间

数据多尺度表达的过程中位置不确定性是趋于正态分布的。因此，可以认为误差椭圆是对矢量数据多尺度表达的位置不确定性可视化的有效方法，具体的用误差椭圆的参数描述在不同位置、不同尺度下的空间数据不确定性，误差描述符的大小取决于空间数据多尺度表达的不确定性大小，如图6.2所示。

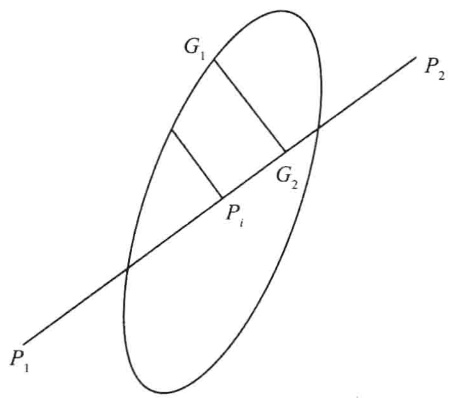


图6.2　误差椭圆

误差椭圆方法不仅具有单独显示某一空间特征位置不确定性的能力，而且还具有同时显示同一幅地图上所有空间目标位置不确定性的能力。后一特性可以提供地图上所有空间特征的总体精度描述，使用户对空间特征的不确定性空间分布有初步的比较。其缺点是：当空间特征及其相应误差椭圆描述符同时显示在一张地图上时，某些空间特征可能会与其他的空间特征误差描述符相互重叠。这一弱点可以通过每次只显示一个或者几个空间特征的误差描述符的方法加以克服。

2. 误差带

运用误差传播律，我们可以将点元的上述误差椭圆概念拓展到描述线元和多边形的位置不确定性描述。误差带模型是线状要素位置不确定性的评价方法中最典型的方法，如"ε-带"、"E-带"、"G-带"、"C-带"、"H-带"等。这些模型都是为评价实测或数字化线的

不确定性发展起来的。如今，关于线状要素的表达已发展到多个尺度(多重表达)，甚至是无级表达。已有的误差带模型能否适用于线状要素的多尺度表达的不确定性评价，关键是看线状要素多尺度表达的不确定性是否满足建立误差带的基本条件。

Chrisman(1982)引入"ε-带"，如图6.3所示，主要是用于评价数字线的线位精度，其基本假设是：GIS中的数字线是将客观实体抽象为真实线后，通过随机扰动生成，线的误差为随机误差；生成数字线的过程是不相关的。对于简单的线元，如图6.3，其"ε-带"由线元端点处以ε为半径的两个半圆和线元中间以2ε为宽度的平行线构成。与线元对应的真实线将以一定的概率落入"ε-带"中。Honeycutt(1986)考虑到线元上点与端点的关系，对"ε-带"进行了扩展，发展了"E-带"。"E-带"假设端点间端点内的误差间是不相关的，端点坐标方差相等，线元上任意点由其端点唯一确定，其点位误差由误差传播定律求得。顺次连接线元上各点的点位误差即为"E-带"。

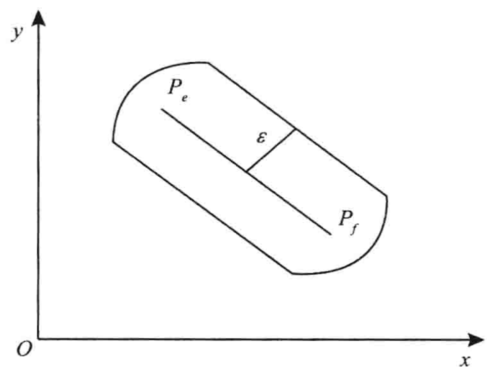


图6.3 "ε-带"(Chrisman，1982)

考虑到了线元端点间的误差是不同精度，甚至是相关的许多情况，在"E-带"的基础上进行了进一步的研究，发展了"G-带"。"G-带"先求出线元上任意一处误差椭圆的长、短半轴和极值的方位，然后将这些误差椭圆簇绘出，构成误差椭圆簇的包络线，这样就构

成了“带”。误差带的边界线方程是由4个部分组成，如图6.4所示。

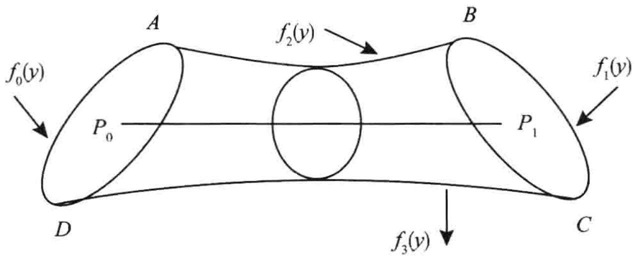


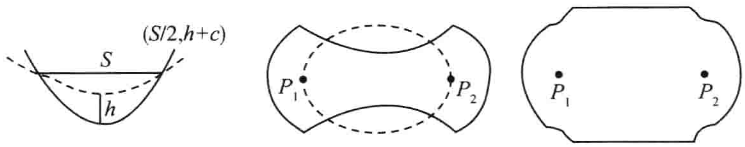
图6.4　“G-带”(史文中等，2000)

“ε-带”、“E-带”、“G-带”都是一定意义上的置信带。范爱民等人(2001)利用信息熵理论，在“ε-带”、“E-带”、“G-带”的基础上发展了“H-带”(误差熵带)。“H-带”不是任何意义上的置信带，而是一种完全确定的、与置信水平无关的不确定带模型。考虑到空间数据多尺度表达对线元位置不确定性的影响，线元数据不确定性与空间数据多尺度表达的不确定性相结合的误差带，如图6.5所示。数据不确定性与空间数据多尺度表达的不确定性的综合影响$\sigma_{综}$为：

$$\sigma_{综}^2 = \sigma_{数}^2 + \sigma_{多}^2 \tag{6.1}$$

式中，$\sigma_{数}$为数据不确定性，且取$\sigma_{数}^2 = \sigma_x^2 + \sigma_y^2$；$\sigma_{多}$为空间数据多尺度表达的不确定性，其不确定性来自于原有线上的误差$\sigma_{原}$。

“ε-带”、“E-带”、“G-带”均是在一定的假设条件下，通过线元端点的误差，利用误差传播定律求出线元上各点的误差或误差椭圆(精度指标)。这三种误差带都是与一定置信水平相关的置信区间。“H-带”是与置信水平无关的不确定带，是“ε-带”、“E-带”、“G-带”的扩展。线元数据不确定性与空间数据多尺度表达的不确定性相结合的误差带同样也是基于误差椭圆的。因此，线元误差带模型的关键是要求出线元上各点的点位误差或误差椭圆。



（a）多尺度表达产生的不确定性（b）不确定性的叠加 （c）不确定性的综合

图 6.5 数据不确定性与空间数据多尺度表达的不确定性的综合

6.3.2 灰度图

灰度图适用于描述某一类具有不确定性的数据，但需要明确给出不确定性的度量方法。史文中(2005)对于栅格数据的不确定性可视化进行了大量的研究，因此本节主要对矢量数据进行研究。在矢量数据中，空间数据多尺度表达的不确定性主要是对点的度量；以此通过"E-带"度量线上的不确定性；多边形的不确定性源于多边形边界的变化，其不确定性的度量类似于线。在空间数据多尺度表达的过程中，其不确定性的范围太小，会产生对比不足的问题，使各个实体的不确定性分辨不清。为此，需要将每个实体的不确定性变换成为一定的灰度范围，并设置灰度变化的规则(包括渐变、突变等变化类型)，从而具体地实现对于不确定性的灰度变量范围，达到图像增强的目的。

设空间数据多尺度表达中的点 (x, y) 处的不确定性可以定量地表达为 $f(x, y)$，通过映射函数 T，对应的灰度级为 $g(x, y)$，即

$$g(x, y) = T[f(x, y)] \tag{6.2}$$

1. 线性灰度变换

将空间数据多尺度表达中的不确定性进行线性变换为灰度值，常能增强可视化的效果。对于多尺度表达中的实体的不确定性范围为 $[a, b]$，变换后实体的灰度值的范围为 $[c, d]$，则有

$$g(x, y) = \frac{(d-c)[f(x, y)-a]}{b-a} + c \tag{6.3}$$

2. 非线性灰度变换

当用某些非线性函数，如对数、指数函数作为映射函数时，可实现不确定性到灰度值的非线性变换。当希望对不确定性的低值区域有较大的扩展，而对高值区域进行压缩时，常用的是对数变换，其一般公式为：

$$g(x, y) = a + \frac{\frac{\ln[f(x, y) + 1]}{b}}{\ln c} \tag{6.4}$$

其中，a、b、c 为可调参数。

当希望对不确定性的低值区域进行压缩，而对高值区域有较大的扩展时，常用的是指数变换，其一般公式为：

$$g(x, y) = b^{c[f(x, y) - a]} - 1 \tag{6.5}$$

3. 灰度分层切片

在可视化表达中，如果需要将 $[e, f]$ 范围抽象出来，转换成最大的灰度值或重点表示出来，则可以表达为：

$$g(x, y) = \begin{cases} \max, & e < f(x, y) < f \\ 0, & \text{其他} \end{cases} \tag{6.6}$$

用灰度图方法表示空间数据多尺度表达中的不确定性，一般的表示方法是：灰度越深表示不确定性越小；灰度越低表示不确定性越大。图6.6是灰度图对空间中线的多尺度表达过程中的不确定性的可视化表达。

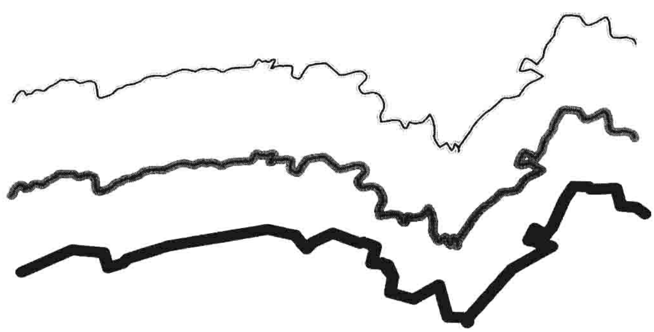


图6.6　线状要素多尺度表达的不确定性可视化（灰度图）

6.3.3　色彩图

色彩是感知差异的一种常见方法，因此是可视化的重要变量之一。计算机制图学已经发展了多种用于上图形的色彩模型。Jiang 等人(1995)、Jiang(1996)提出了不确定性可视化的 HIS 模型；史文中(2005)提出了不确定性可视化的 RGB 色彩模型。这些色彩模型在不确定性可视化方面具有较强的实用价值，但其关注的主要是栅格数据，本节主要对矢量数据的不确定性的色彩图方法进行讨论。类同于灰度图的方法，将空间数据多尺度表达的不确定性与 RGB 空间的一个点或 HIS 空间中的一个点建立起对应的关系；然后用色彩图来表达不确定性的每个格网单元。色彩图在表达空间数据多尺度表达的不确定性的时候也容易出现描述不清的问题，因此在必要的时候色彩图方法应当结合其他的可视化变量以有效地描述空间实体的不确定性。

1. 空间数据多尺度表达不确定性的分级模型

在空间数据多尺度表达的过程中，先是对不确定性建立分级模型，之后采用色彩图法进行可视化表达，主要是为了反映不确定性传播的规律性和差异性。从视觉角度来说，人们能够辨识的等级差别是有限的。通过限制分级数在一定程度上可以增强可视化的效果，但为了满足可视化表达精度的要求，又要保留足够的分级数。在空间数据多尺度表达的不确定性分级模型中主要采用等差分级模型、等比分级模型和统计分级模型。

①等差分级模型：是不确定性分级的重要方法，主要有界限等差分级模型和间隔递增等差分级模型。

②等比分级模型：在不确定性分级中应用较少，主要有界限等比分级模型和间隔等比分级模型。

③统计分级模型：在统计分级模型中，分级界限的确定是以一些统计量为基础的，这类分级模型确定的分级界限能够较好地反映数据的分布特征。主要有面积相等分级模型和正态分布分级模型。

2. 基于色彩图的不确定性分级表达

(1)RGB 色彩模型

为了避免RGB模型中可能产生的严重问题(Hengl，2004)，可以通过定义一种规则，分别表示不确定性的大小，如类似于地形图中的海拔，从蓝色、绿色、棕色到紫色的变化；一般来说用空间数据多尺度表达不确定性的分级模型所划分的类别是有限的，在用RGB色彩模型进行不确定性信息的可视化时，通常采用加色法。

三原色及其混合颜色中的一种颜色的范围值为[0，255]，则可以通过不确定性的分级模型划分的类别，在其中选择级差最大的对应的颜色种类，与其对应表示。类似于灰度图法，设空间数据多尺度表达中的点(x, y)处的不确定性可以定量表达为$f(x, y)$，通过映射函数T，对应的灰度级为$g(x, y)$，即

$$g(x, y) = T[f(x, y)] \tag{6.7}$$

同样可以采用线性变换、非线性变换等方式建立不确定性的度量值与RGB色彩模型之间的对应关系，对空间数据多尺度表达的过程中的不确定性进行可视化表达。图6.7是RGB色彩模型对空间中线的多尺度表达过程中的不确定性的可视化表达。

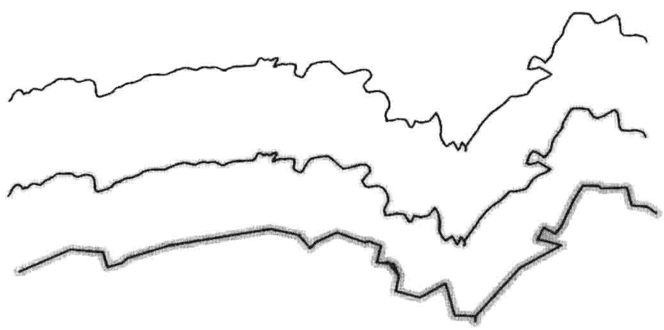


图6.7　RGB色彩模型对线状要素多尺度表达的不确定性的可视化表达

(2)HIS色彩模型

HIS色彩模型的颜色参数为色调(H)、饱和度(S)以及亮度(I)。HIS是以三维空间中的六棱锥来定义的。该六棱锥顶部为一个六边形，其边界线代表不同的色调。六边形中的水平与竖向轴线分别代表饱和度和亮度。Hengl等人(2004)给出的用于色彩模型与

RGB 模型之间的变换关系：

$$\begin{cases} H=\frac{360}{2\pi}\arctan\left(\frac{\sqrt{3}}{2}[G-B],\ R-\frac{G+B}{2}\right)\times\frac{240}{360} \\ S=\sqrt{R^2+G^2+B^2-RG-RB-GB}\times\frac{240}{255} \\ I=\frac{R+G+B}{3}\times\frac{240}{255} \end{cases} \tag{6.8}$$

其中，输入值 R，G 和 B 的变化范围为[0，255]，而输出值 H，S 和 I 在[0，240]范围内变动。可以用 HIS 色彩模型间接实现空间数据多尺度表达的不确定可视化表达。

6.3.4 地图符号法

借助于图形元素变化与组合效果，表象性符号能够成为空间数据表达多尺度的不确定性的可视化方法。这种能够引起视觉差别的图形和色彩变化因素称为“视觉变量”。有了它，地图符号就具有描述不确定性的功能。

1. 地图符号法的设计要求

为了准确地描述空间数据多尺度表达的不确定性，地图符号的设计应具有一定的功能和基本条件满足不确定性表达的需求。

①图案化。图案化就是对要表达的不确定性进行整理、夸张、变形，使之成为比较简单的规则化图形。

②清晰性。符号清晰是不确定性可视化的基本条件，每个符号都应具有良好的视觉特性，要从简单性、对比度和紧凑型等三个方面确定不确定性可视化的符号。

③系统性。系统性是指符号群体内部具有一定的逻辑性，不能孤立地设计每个符号，这是符号能够配合使用表达不确定性的必要条件。在符号形式上表现出不确定性的分类、等级、主次和虚实等关系。

④适应性。用设计符号来描述不确定性数据时，应当考虑空间数据多尺度表达所采用的方法和过程，以便使被选符号具有一定的适应能力。此外，针对不同的可视化用途和不同的用户需求，应设

计不同类型的符号。

2. 地图符号法的表示方法

通过基本图形元素变化与组合，表象性符号能形成众多类型和形式，形成视觉或图形变量系统，可以用来描述空间数据多尺度表达的不确定性的性质和特征。法国人 J. Bertin(1985)领导的巴黎大学图形实验室总结出一套图形符号的变化规律，提出了包括形状、方向、尺寸、明度、密度和颜色的视觉变量。从不确定性表达的角度来看，视觉变量应包括形状、尺寸、方向、明度、密度、结构、颜色及位置。通过这些视觉变量可以将不确定性分布进行编码并表达在地图上，其中位置是符号本身的固有属性(史文中，2005)。在所有基于符号的不确定性可视化方法中，采用的符号都是基于形状、尺寸大小、颜色、方向、纹理以及位置而定义的。其中，在 J. Bertin(1985)给出的图形变量中，尺寸是确定图中维数大小量值的一种表达方式。

在空间数据多尺度表达的不确定性可视化中，要表达定位特征、性质特征、数量特征和不同尺度特征。在这些特征中定位特征是空间对象的基本标志之一，不需要考虑；空间结构特征难以表达。因此，空间数据多尺度表达的不确定性可视化中主要考虑性质特征、数量特征和尺度特征。

①性质特征。在空间数据多尺度表达过程中，描述不确定性的性质特征的变量主要是形状、方向和颜色，也可以辅助用明度、密度等变量起到增强的作用。

②数量特征。空间数据多尺度表达的不确定性需要用什么类型的符号来表示，不仅仅与其本身的性质、可视化的要求有关，更与不确定性的数量特征有关。空间数据多尺度表达可以用“数值处理”和“分级处理”两种形式实现。前者属于精确的描述方法，后者则是相对概略的描述方法。空间数据多尺度表达的不确定性用地图符号法来表示，主要使用的是“分级处理”方式。所谓的分级处理方式就是在视觉模拟上使用分级符号。分级符号在视觉变量的选择上要突出等级感，然后用文字对各等级的符号赋予相应的不确定性数据范围。因此，分级符号的数量是由等级感转换而来的。表达空

间数据多尺度表达不确定性的数量特征的变量相对较少，尺寸是表现准确数值关系的有效方式。表现数量相对大小的顺序或等级既可用尺寸，也可用明度、结构等变量，如图 6.8 所示。

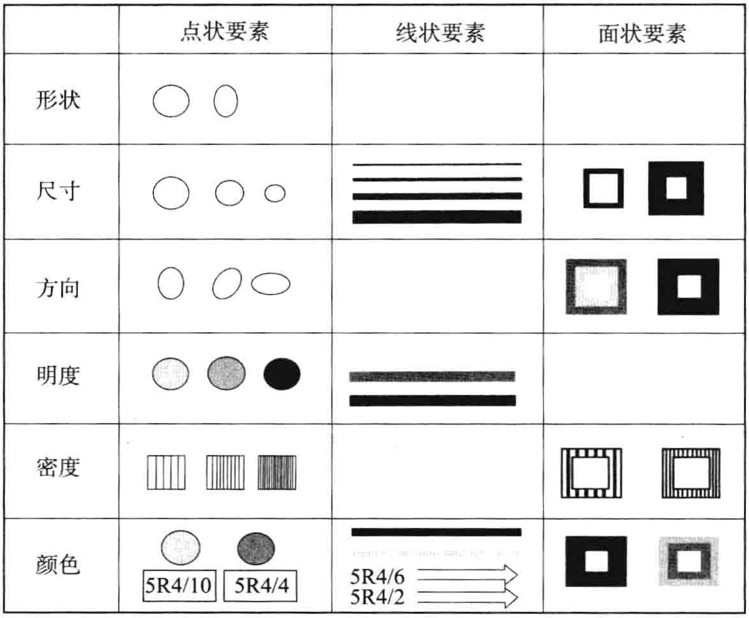


图 6.8　空间数据多尺度表达的不确定性可视化视觉变量

③尺度特征。尺度特征就是空间数据的表达的尺度标志，能够反映不同尺度下的不确定性的产生、传播和累积特点。描述不确定性的尺度特征的变量主要是形状和颜色，也可以辅助用明度、密度等变量起到增强的作用。时间特征很难用静态的二维图形直接表达，但可以用不同尺度下的地图符号进行标示并配以文字加以说明。

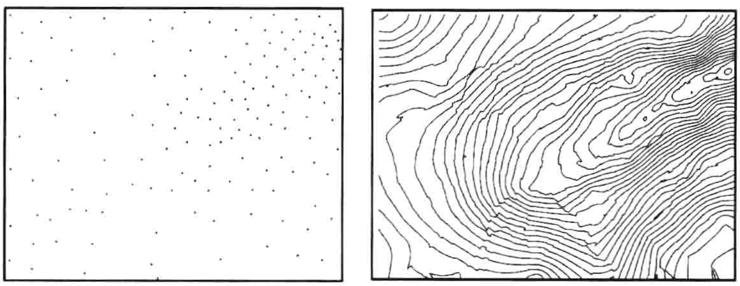
6.4　三维可视化

随着计算机技术和信息技术的快速发展，三维可视化日益受到

重视，并成为一个热门研究方向。Foley等人(1994)对三维不确定性描述方法及可视化表达进行了研究，史文中(2005)对空间数据的三维可视化方法进行了研究。本节重点研究如何用三维可视化的技术实现空间数据多尺度表达的不确定性模型。

6.4.1　等值线法

等值线法是用一组等值线来表示连续面状分布的制图现象数量特征渐变的方法。用等值线法对空间数据多尺度表达的不确定性进行可视化表达，其不确定性数据被作为数据集的Z值来对待。等值线的绘制方法较多，可以分为规则格网法和不规则格网法。其关键是在构建格网的基础上，寻找各格网边上的等值点，并对等值点进行连接和光滑。在此基础上还可以综合等值线注记、色彩等方面，表达出空间数据多尺度表达区域内的不确定性变化及分布状况。图6.9中的(a)图是空间中的离散点，这些离散点都具有自身的不确定性，(b)图是根据其不确定性绘出的不确定性等值线。



(a)　　　(b)

图6.9　等值线法表示空间数据多尺度表达的不确定性

其中，等值线的值越大表示该区域的不确定性值越大。用等值线法表示空间数据多尺度表达不确定性的主要特征是直观形象、并容易与二维可视化方法配合使用，但是在表达三维方面具有较大的局限性。

6.4.2 基于 DTM 的不确定性可视化拓展

1. 基于 DTM 的矢量数据三维可视化表达方法

对于以矢量数据或以矢量形式表达的空间数据结果而言，在三维情况下，误差椭圆模型发展为误差椭球模型，该误差椭球模型主要用椭球描述三维点元的位置不确定性，如图 6.10 所示。二维情形下的某些不确定性度量指标可以以标量形式来表达。为了表达不确定性的空间分布，史文中(2005)用三维柱状图来表示空间某一点位置的不确定性，立柱越高表示该点的不确定性越大。图 6.11 给出了某点状图不确定性指标可视化的实例。

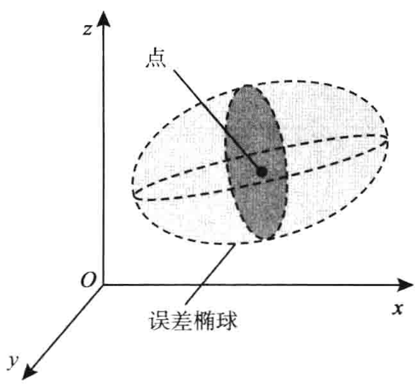


图 6.10　三维情形中空间点元的误差椭球不确定性模型(史文中，2005)

在矢量数据中，基于误差传播律可以估计出空间数据表达中的离散点的不确定性，并以该值作为 DTM 的 Z 值，通过利用不规则三角网 DTM 直接进行不确定表面的构建，将不确定性表面剖分为连续的相互连接的三角面，三角面的形状和大小取决于离散点的密度和位置。此方法的关键就在于不规则三角网的构建。

构建不规则三角网，就是由已知的离散数据点的空间位置构建三角网，即每三个离散数据点构成一个三角形，也称为自动连接三角网。对于平面上 n 个离散点，将其中相近的三个点构成最佳三角形，使离散点都成为三角形的顶点，如图 6.12 所示。构建三角网

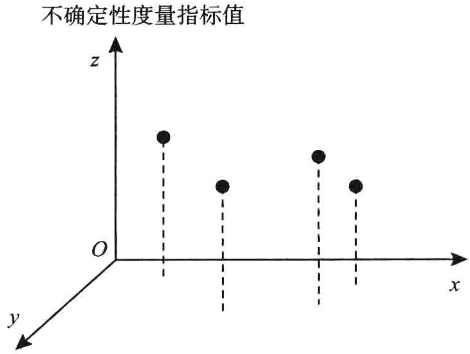


图 6.11　三维情形中具有不确定性信息的柱状图(史文中，2005)

的方法有很多，最常用的是 Tsai(1993)提出的构造 Delaunay 三角网的凸包插值算法。

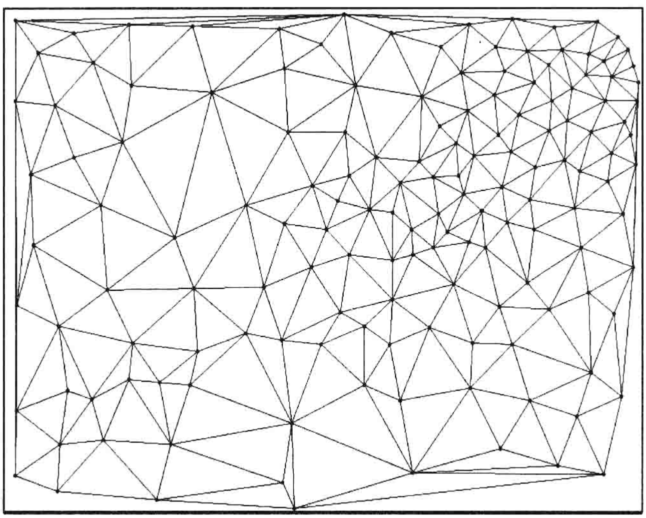


图 6.12　离散点构造的 Delaunay 三角网

对于矢量数据来说，基于 DTM 的多尺度表达的三维可视化表达方法的优点是：能较好地表达一定范围内的复杂的不确定性，应

用也较为方便，但是也存在数据结构复杂、不便于规范化管理等缺点。

2. 基于 DTM 的栅格数据三维可视化表达方法

类似于矢量数据基于误差传播律可以估计出栅格地图或遥感影像表达中的离散点的不确定性。栅格数据的多尺度表达不确定性被认为是 DTM 中的 Z 值。为了将不确定性表达在一个光滑的三维表面上，可以通过在栅格数据中内插不确定性数据来实现。由于多尺度表达中的不确定性是离散的，因此每个栅格单元的 Z 值的获取是基于 DTM 的栅格数据的三维可视化表达方法的关键。最直接的解决办法是通过离散数据的内插产生每个栅格单元的不确定性(误差)。栅格图上每个栅格所处位置的高度越高，意味着该处不确定性越大。

构建规则格网 DTM 是因为基于离散点的多尺度表达的不确定性也是呈离散分布的。其基本思路是：选择适当的数学函数模型，利用已知栅格上的不确定性求出函数的待定系数，然后求算每个栅格上的不确定性。通常使用距离加权平均法和多项式内插法等方法构建规则格网 DTM。内插方法对 DTM 精度的影响取决于原始采样点的密度和分布。

(1)距离加权平均法

设所需计算栅格 p 的平面坐标为（x_p，y_p），多尺度表达过程中的不确定性为 z_p，则

$$z_p = \sum_{i=1}^{n} w_i \cdot z_i / \sum_{i=1}^{n} w_i \tag{6.9}$$

式中，$i = 1, 2, \cdots, n$ 表示取 p 点周围的 n 个离散点；z_i 为第 i 点的不确定性；w_i 为第 i 点的权函数，

$$w_i = 1/d_i^t \tag{6.10}$$

令 $t = 2$，则 i 点（x_i，y_i）到格网点（x_p，y_p）的距离为：

$$d_i = \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2} \tag{6.11}$$

在构建格网的过程中，经常需要选取与插值点距离最近的 n（通常取 4 ~ 12）个离散点的不确定性进行控制。以 p 点为圆心，r 为半径进行搜索。半径 r 可适当变化，使圆内离散点数控制在需求

数目即可。初始半径的计算公式为：

$$r = \sqrt{\frac{7S}{\pi N}} \tag{6.12}$$

其中，S 为多尺度表达区域的面积，N 为多尺度表达区域内总的离散点数。

为了加快搜索速度，也可以在插值点 p 周围构建正方形的选取框，判断落入框内的离散点的个数。当落入框内的离散点较多时，可缩小选取框的尺寸；反之，增大选取框的尺寸。选取框初始边长的计算公式为：

$$L = \sqrt{\frac{nS}{N}} \tag{6.13}$$

其中，S 为多尺度表达区域的面积，N 为多尺度表达区域内总的离散点数。

(2)多项式内插法

线性内插法和双线性多项式内插法是多项式内插法中较为简单实用的两种方法。其数学模型分别为：

$$z = a_0 + a_1x + a_2y \tag{6.14}$$

$$z = a_0 + a_1x + a_2y + a_3xy \tag{6.15}$$

其中，z 为多尺度表达过程中的不确定性；a_0，a_1，a_2，a_3 为待定系数。

在内插的过程中，将与 p 距离最近的 3 ~ 6 个点的坐标值和高程值代入方程，使用最小二乘法即可解出全部系数，然后将插值点的坐标代入方程，即可计算出该点多尺度表达过程中的不确定性。选取数据点时，可采用式(6.13)中的方法进行搜索。

采用离散的实验数据，对其不确定性进行计算，并通过上述内插方法，获取的规则格网的不确定性的三维可视化表达如图 6.13 所示。

应用规则格网 DTM 建立的空间数据多尺度表达中的不确定性可视化模型具有结构简单、应用方便和易于管理的优点，但也存在精度较低、存在数据冗余或难于表现局部不确定性等缺点。

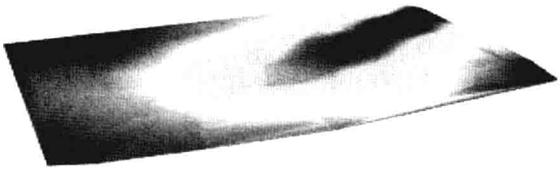


图6.13 基于规则格网的不确定性的三维可视化

6.5 动态可视化

伴随着计算机技术和信息技术的发展，动态可视化技术应运而生。动态可视化是科学可视化的重要内容，也是近年来的研究热点问题。采用动态可视化的形式在地图、GIS 和遥感中已有应用(祝国瑞，2003；史文中，2005)。空间数据多尺度表达的不确定性也可以通过动态的形式来表达。动态可视化的方法以简明、快速而连续的方式集中地显示出由于空间数据多尺度表达过程中出现的不确定性。本节将主要探讨动态符号和动画两个方面的问题。

6.5.1 动态可视化的特征和作用

空间数据多尺度表达的不确定性动态可视化是采用动态符号和动态地图以及计算机图形学理论，通过计算机动画、计算机高级显示技术，把空间数据多尺度表达过程中的不确定性用动态的方式在一定的媒介(如电子屏幕)上表达的过程。其最主要的特征是逼真而形象地表现出空间数据多尺度表达过程中的不确定性过程和规律，也就是变化过程中的特点。具体来说，可以用于表示以下内容：

①动态模拟。模拟使空间数据多尺度表达过程中的不确定性变化过程再现，在这些复杂的动态表达过程中，动态可视化是个得力的帮手。

②运动模拟。通过增加或降低可视化的速度，来对变化过程做

一个动态模拟，在需要的时候还可以暂停可视化的过程以仔细检查某一尺度下的不确定性。

③多视角显示。通过改变视角可以获取多尺度表达过程中不同视角下的不确定性信息。

6.5.2 动态可视化的表示方法

动态可视化可以采用各种方法及组合来表示多尺度表达过程中的不确定性信息。常用的表示方法有：

1. 动态符号法

采用传统地图学理论中的视觉变量组成动态符号，或使用专题地图制图的运动线法来表示空间数据多尺度表达过程中的不确定性信息。也可以采用动态视觉变量来表达空间数据多尺度表达中空间实体的整个生命周期(艾廷华，2009)的不确定性变化时长、速率、次序及节奏。因此，需要设计相应一组动态符号，并加上相应表达手段，如闪烁、跳跃、色度、亮度等变化来反映空间实体不确定性变化的特征。动态符号法可以通过传统的视觉变量来表达多尺度表达中的不确定性，还可以表达时间变量，通过设置符号在不同尺度下的颜色值，以及设置颜色变化的规则(包括渐变、突变等变化类型)，来具体地实现对于某一单个符号在颜色变量方面的动态变化。因此，动态视觉变量可以用图 6.14 来表达。

2. 连续快照法

采用连续快照的方法将不同尺度下的不确定性以一组图形的形式展现出来。这是采用一系列状态对应的图形来表现不同尺度下的不确定性信息。这一方法对不同尺度下的不确定性表现方面是较为全面的，但对不确定性的变化过程的表达不够明确，同时为了实现连续快照，存在大量的冗余数据。

3. 动画法

动画方法可以显示空间数据多尺度表达的不确定性。其制作方法上是通过适当地在表达中内插足够密度的快照，使多尺度表达中的不确定性的状态差异由突变变为渐变。这一方法弥补了连续快照法中变化不够明确的问题，是动态可视化中表现较为丰富的形式。

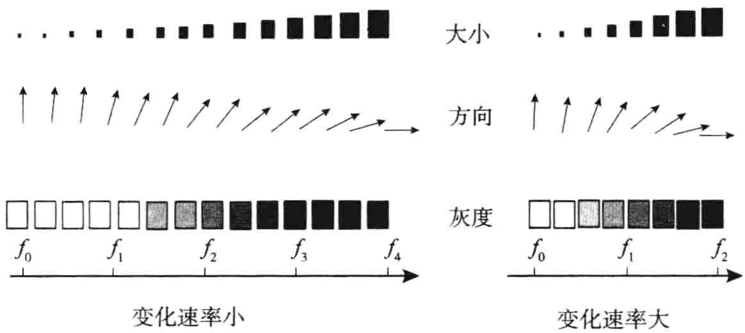


图6.14 空间数据多尺度表达的不确定性可视化动态视觉变量

本部分主要讨论的是用动画表达多尺度表达过程中空间数据的精度变化。

在矢量数据多尺度表达过程中，空间数据的不确定性是最为重要的内容。利用动画对空间数据的不确定性进行可视化，需得到两个随机指标(史文中，2005)：一是位置不确定性分布，用于描述空间数据多尺度表达过程中空间实体的位置；二是空间特征的随机化(Fisher，1994a，1996b)。在通过不确定性分析后获得的结果可以用动画的方式表达出来。在确定速度和精度的关系的情况下，将不同尺度下的不确定性连续地播放出来(Blenkinsop et al.，2000)，在播放的过程中，如果动画变化较小，则可以认为空间数据的精度较高，并且在多尺度表达的过程中不确定性的变化较小，误差累积较小，较好地保持了空间数据的精度。否则，认为在多尺度表达的过程中误差累积较大、精度随着尺度的变化进一步降低。

6.5.3 动态可视化的设计

不同于一般空间数据多尺度表达的不确定性的可视化，动态可视化设计在可视化过程中是个重要问题，应着重考虑以下几点：

第一，明确了解空间数据多尺度表达的不确定性及其对可视化的要求，了解它所表示的内容和主要目的，确定具体关注的变化是不确定性变化的过程还是最后状态。

第二，拟定表达方法和设计动态符号。对于侧重表现空间数据多尺度表达中不确定性最后状态的，采用动态符号法或动线法设计动态符号；对于侧重于表现变化过程的则要综合采用计算机地图动画、动态符号、闪烁、漫游、连续快照法或其他方法进行设计。

第三，动画或符号设计。动画或符号设计在表达补缺型要素上具有全面、形象、明确的特点，但是其制作是比较困难的，尽管已有相当多的商品化软件，如 Edge Animate、IDL6. X 等，但是还需要根据空间数据多尺度表达的不确定性特点，必须进行精心的设计和制作。

第四，根据测试及用户反馈，反复修改直到运行无误，最后将调试好的文件发布为适用于动态可视化格式的文件。

总之，动态可视化应用与空间数据多尺度表达的不确定性是一种直观、有效的方法，可以表现不确定性产生、传播的内在本质和变化规律。

6.6　多媒体可视化

6.6.1　多媒体的概念

媒体(media)是人与人之间实现信息交流的中介，简单地说，就是信息的载体，也称为媒介。“多媒体”一般理解为“多种媒体的综合”，也就是直接作用于人感官的文字、图形、图像、动画、声音和视频等各种媒体的统称，即多种信息载体的表现形式和传递方式。因此，我们可以把这一概念分成广义和狭义两种。广义的多媒体是指能传播文字、声音、图形、图像、动画和电视等多种类型信息的手段、方式或载体，包括电影、电视、CD-ROM(compact disc read-only memory)、VCD、DVD(digital versatile disc)、电脑、网络等；狭义的多媒体是指融合两种以上“传播手段、方式或载体”的人机交互式信息交流和传播的媒体，或者说是指在计算机控制下把文字、声音、图形、图像、动画和电视等多种类型的信息，混合在

一起交流传播的手段、方式或载体。

6.6.2 基于多媒体技术的空间数据多尺度表达不确定性可视化

空间数据多尺度表达的不确定性可以通过多媒体进行可视化，多媒体可视化的特性包括：多样性、交互性、集成性和超链性(陈棉，2007)，它也不同于一般的 GIS 工具软件，重点在于空间数据多尺度表达中的不确定性的内容查询、浏览和展示等，侧重于不确定性的传输、接受和理解。基于多媒体技术的空间数据表达的不确定性可视化在大多数情况下作为不确定性分析及其表达的辅助型手段，其主要的形式为声音、文字、图形和图像的结合。

1. 声音

声音是数字世界中表现和交流地理数据的一种非视觉工具(史文中，2005)。但是在空间数据多尺度表达的不确定性可视化中，声音是不能单独使用的，必须借助于其他的可视化形式，将声音融入到可视化中。因此，声音是可视化方法的一种补充形式，用于辅助或加强可视化的效果。然而，在 GIS 中，声音并不能单独应用，因为要对感兴趣的地区作出分析，我们必须借助可视化的地图，声音必须融入地图可视化方法之中，才能很好地表达不确定性。因此，声音是对可视化方法的一种补充。例如，在多变量数据情况下，可视化方法在叠置分析的合成图建模方面可能存在困难。而声音则为交互式的可视化显示提供了一种附加的非视觉方法。

通过定义某些抽象声音变量(Krygier，1994)与空间数据多尺度表达的不确定性级别之间的线性关系，实现用声音来描述地图上数据的不确定性。在不确定性可视化问题中，声音还可以作为一种警示器(史文中，2005)。当空间数据多尺度表达中的某个要素的不确定性超过给定阈值时，警报声将会自动发出以唤起用户的注意。根据 Fisher(1994)的研究成果，对空间数据多尺度表达过程中用声音来增强不确定性可视化，通过交互式的方式使用声音来表达其中的不确定性，在此过程中可以根据一定的标准利用不同的音调定义不确定性的大小。

2. 文字、图形、图像等

文字、图形、图像等形式是多媒体可视化的重要组成部分。在空间数据多尺度表达的过程中产生的不确定性可以借助这一形式进行表达。可以通过交互的方法将用户感兴趣的点或区域的不确定性通过文字、图形或图像等形式在多媒体系统中表达出来，对于增强不确定性信息的传输，更有助于用户对不确定性信息的接受和理解。

基于多媒体技术的空间数据多尺度表达不确定性可视化具有以下特点：内容的丰富性、艺术科学性、生动直观性、集成性、查询和浏览的便捷性，这也是衡量空间数据多尺度表达的不确定性可视化的关键。

6.7　本章小结

不确定性可视化是不确定性数据和分析结果的一种表现形式，已成为空间信息阅读、理解进而交互作用中最重要的工具和手段。本章研究了空间数据多尺度表达不确定性的可视化方法。具体以视觉与认知理论为基础，设计了空间数据多尺度表达不确定性可视化流程；在二维可视化方面，提出了空间数据多尺度表达不确定性和空间数据的综合误差带可视化方法，基于灰度图、色彩图和地图符号的空间数据多尺度表达中的不确定性可视化方法；系统研究了空间数据多尺度表达方法的三维可视化方法，主要包括等值线法、基于 DTM 的空间数据多尺度表达不确定性可视化方法，并建立了相关的模型；最后研究了空间数据多尺度表达的动态可视化和多媒体可视化方法。

第7章 结　　语

随着GIS应用领域的不断扩展和需求层次的日益提高，人们越来越多地需要对多尺度的空间数据进行分析、处理和表达。因此，对空间数据的多尺度表达不确定性问题进行研究成为必然。国内外早已对空间数据的不确定性问题进行研究，以空间数据不确定性为主题的国际空间数据质量大会(The International Symposium on Spatial Data Quality)已经召开了8届；两年一届的自然资源与环境科学中的空间精度评价国际会议(The International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Science)已经召开了10届。第八届(2013)国际空间数据质量大会上有数篇空间数据多尺度表达不确定性相关的论文参与交流，在自动综合算法的质量评价中也都包含了对综合结果不确定性的评估，而且许多空间数据不确定性的研究与空间数据多尺度表达不确定性的研究相关，其成果可为空间数据多尺度表达不确定性研究利用和借鉴。但是这些有关空间数据多尺度表达不确定性的研究成果并不多，尚未建立科学合理的研究体系和研究框架。因此，本书对空间数据多尺度表达的不确定性及其评价指标、量化模型，空间数据多尺度表达空间关系不确定性及其评价指标、量化模型，空间数据多尺度表达不确定性可视化方法等进行了系统深入的研究。本书的主要内容包括以下七个方面：

第一，分析了与空间数据多尺度表达的不确定性相关的基本概念，包括空间数据多尺度表达、不确定性、空间数据的不确定性、空间数据多尺度表达的不确定性等，在本书中将空间数据多尺度表达的不确定性定义为空间数据多种比例尺表达的空间数据与空间现象本身固有的属性之间的差异。

第二，讨论空间数据多尺度表达不确定性的研究意义及相关研究进展。对空间数据多尺度表达的不确定性研究在空间数据多尺度表达不确定性认知、多尺度产品质量评价与控制、决策分析中都有非常重要的意义。

第三，分析空间数据多尺度表达不确定性模型的数学基础，论述和分析已有的空间数据随机不确定性和模糊不确定性模型。处理现实世界随机不确定性和模糊不确定性所涉及的几种实用而且常见的数学理论包括概率论与数理统计、信息论与熵、模糊理论、粗集理论。随机不确定性模型包括各种误差带理论，模糊不确定性模型包括基于模糊度的不确定地理目标的形式化描述方法和不确定地理目标的粗集表达方法。这些数学理论与不确定性模型是构成本书研究空间数据多尺度表达不确定性的理论基础。

第四，建立空间数据多尺度表达不确定性的形式化表达方法。该形式化表达方法是在充分研究空间数据多尺度表达的实现方案、空间数据多尺度表达不确定性来源和类型的基础上构建的。

第五，构建空间数据多尺度表达位置不确定性的评估模型。根据空间数据多尺度表达的生成机制以及不确定性在这个过程中的传播机理，构建空间数据多尺度表达不确定性的评价指标和量化模型。重点构建了点元、线状要素、面状要素、点群、居民地多尺度表达的不确定性的评价指标和量化模型。

第六，构建空间数据多尺度表达空间拓扑和方向关系不确定量的评价模型。该模型是在充分分析空间数据多尺度表达空间拓扑和方向关系的不确定性的基础上，利用粗集理论进行构建的。

第七，研究空间数据多尺度表达不确定性的可视化方法。本书以视觉与认知理论为基础，设计了空间数据多尺度表达不确定性可视化流程，并重点研究了二维可视化、三维可视化、动态可视化、多媒体可视化在空间数据多尺度表达不确定性的可视化中的应用方法。

主要参考文献

[1] ABBAS I, GRUSSENMEYER P, HOTTIER P. Contrôle de la planimétrie d'une base de données vectorielle: une nouvelle méthode basée sur la distance de Hausdorff: la méthode du contrôle linéaire [J]. Bulletin SFPT, 1995, 137 (1):6-11.

[2] ABDELMOTY A I, WILLIAMS M H. Approaches to the representation of qualitative spatial relations for geographic databases[J]. Geodesy, 1994, 40(1):204-216.

[3] ALLEN J. Maintaining knowledge about temporal intervals [J]. CACM, 1983,26(11): 832-843.

[4] ANTONIO T MOZAS, FRANCISCO J ARIZA . New method for positional quality control in cartography based on lines: a comparative study of methodologies [J]. International Journal of Geographical Informa-tion Science, Vol. 25, Nos. 10-12, October-December 2011: 1681-1695.

[5] A U FRANK B. Qualitative spatial reasoning about distances and directions in geographic space [J]. Journal of Visual Languages and Computing, 1992, 3(2): 343-371.

[6] AI TINGHUA, LI ZHILIN, LIU YAOLIN. Progressive transmission of vector data based on changes accumulation model[C]. The 11th International Symposium on Spatial Data Handling, SDH 2004, Berlin,2004:459-467.

[7] ALBER R F. The national science foundation national center for geographic information and analysis [J]. International Journal of Geographical Information Systems, 1987, 3(1):117-136.

[8]BALLEY S, PARENT C, SPACCAPIETRA S. Modelling geographic data with multiple representations [J]. International Journal of Geographical Information Science, 2004, 18(4): 327-352.

[9]BERTIN J . Graphical semiology [M]. Madison Wisconsin: University of Wisconsin Press, 1985.

[10]BLAIS J A. On some practical applications of information theory in geomatics [J]. Candian Institute of Surveying and Mapping (CISM) Journal, 1991, 45(2):239-247.

[11]BLENKINSOP S, FISHER P E, BASTIN L, et al. Evaluating the perception of uncertainty in alternative visualization strategies [J]. Cartographica, 2000,37:1-14.

[12]BURROUGH P A. Principles of geographical information systems for land resources assessment [M]. Oxford: Clarendon Press, 1986.

[13]BUTTENFIELD B P. A rule for describing line feature geometry [C]// MCMASTER R B, BUTTERFIELD B P. Map Generalization: Making Rules for Knowledge Representation, Longman, Harlow. 1991:150-71.

[14]BRUGGER B P. Spatial Theory for the integration of resolution-limited data [D]. Maine:University of Maine, 1994.

[15]CASPRY W, SCHEURING R. Error-band as measurers of geographic accuracy [C]// EGIS Foundation. Proceeding of EGIS'92. 1992: 226-233.

[16]CASPRY W, SCHEURING R. Positional accuracy in spatial databases [J]. Computer Environ and Urban System, 1993, 17 (2):103-110.

[17]CECCONI A. Integration of cartographic generalization and multi-scale databases for enhanced web mapping [D]. Zurich: University of Zurich, 2003.

[18]CHANG S, SHI Q, YAN C. Lconic indexing by 2-D strings [J]. IEEE trans on pattern analysis and machine intelligence, 1987,9 (3):413-428.

[19] CHEN XIAO-YONG. Spatial relations of distance between arbitrary objects in 2D/3D geographic spaces based on the hausdorff metric [C]. LIESMARS'95. Wuhan, 1995.

[20] CHEUNG C K, SHI W Z. Estimation of the positional uncertainty in line simplification in GIS [J]. The Cartographic Journal, 2004, 41(1):37-45.

[21] CHUI KWAN CHEUNG, WENZHONG SHI. Estimation of the positional uncertainty in line simplification in GIS [J]. The Cartographic Journal, 2004, 41(1): 37-45.

[22] CICERONE S, DI FELICE P. Cardinal relations between regions with a broad boundary [C]. 8th ACM Symposium on Advances in Geographic Information Systems (GIS' 00). New York: ACM Press, 2000.

[23] CHENG TAO, MOLENAAR M, LIN HUI. Formalizing fuzzy objects from uncertain classification results [J]. International Journal of Geographical Information Science, 2001, 15(1):27-42.

[24] CLEMENTINI E, FELICE P D, OOSTEROM P V. A small set of format topological relationships suitable for end-user interaction [C]. Advances in Spatial Databases. 1993: 227-295.

[25] CLARKE B L. A calculus of individuals based on connection [J]. Norte Dame Journal of Formal Logie, 1981, 2(3):287-296.

[26] COHN A G, GOTTS N M. The "egg-yolk" representation of regions with indeterminate boundaries [C]// BURROUGH P A, FRANK A U. Geographic Objects with Indeterminate Boundaries. London: Taylor & Francisco, 1996: 171-187.

[27] COHN A G. The challenge of qualitative spatial reasoning [J]. ACM Computer surveying, 1995, 27(3):323-325.

[28] DRIK BURGHARDT. Controlled line smoothing by snakes [J]. Geoinformatica, 2005, 9(3):237-252.

[29] EGENHOFER M J, Mark D. Nalve Geography[C]// FRANK A, KUHN W. Lecture Notes in Computer Science 98: Conference on

Spatial Information Theory: A Theoretical Basis for GIS. Berlin: Springer-Verlag, 1995: 1-5.

[30] EGENHOFER M, CLEMENTINI E, DI FELICE P. Evaluating Inconsistencies Among Multiple Representations [C]. Sixth International Symposium on spatial Data Handling, Edinburgh. Scotland, 1994.

[31] EGENHOFER M, FRANZOSA R D. Point-Set topological spatial relations [J]. International Journal of Geographical information systems, 1991,5(2):161-174.

[32] FISHER P. Hearing the reliability in classified remotely sensed images [J]. Cartography and Geographical Information Systems, 1994,21:31-36.

[33] GALTON A. The Mereotopology of Discrete Space [C]. Conference on Spatval Information Theory. COSIT'99, Springer-Verlag Berlin Heidelberg, 1999: 251-266.

[34] GOODCHILD M, HUNTER G. A simple positional accuracy for linear features[J]. International Journal Geographical Information Science, 1997,11 (3): 299-306.

[35] GOODCHILD M, BUTTENFIELD B, WOOD J. Intro-duction to visualizing data validity [C]// HEARNSHAW H M, UNWIN D J, CHICHESTER, WEST SUSSEX. Visualization in Geographical Information Systems. New York: Wiley and Sons,1994: 141-149.

[36] GOODCHILD M, CHANG C L, LEUNG Y. Visualizing fuzzy map [C]// HEARNSHAW H M, UNWIN D J, CHICHESTER, WEST SUSSEX. Visualization in Geographical Information Systems. New York: Wiley and Sons, 1994: 158-167.

[37] GOODCHILD M, QUATTROCHID A. Scale, multiscaling, remote sensing and GIS [C]// QUATTROCHID A, GOODCHILD M. Scale in Remote Sensing and GIS. Boca Rotan: Lewis Publishers, 1997: 1-11.

[38] GOODCHILD M F. The issue of accuracy in spatial databases

[C]// MOUNSEY H, TOMLINSON R. Building databases for global science. London: Taylor & Francis, 1988:31-38.

[39] GOYAL R K, EGENHOFER M J. Similarity of direction relations [C]. Seventh International Symposium on Spatial and Temporal Databases. Los Angeles, California, USA, 2001.

[40] HERZOG A. 1989. Modelling reliability on statistical surfaces by polygon filtering[C]// GOODCHILD M, GOPAL S. Accuracy of Spatial databases. New York: Taylor and Francis. 209-218.

[41] H TODD MOWRER, RUSSELL G CONGALTON. Quantifying spatial uncertainty in nature resource: Theory and Application for GIS and Remote Sensing [M]. Chelsea: Sleeping Beer Press, 2000.

[42] HABIB, SCHENK. A new approach for matching surfaces from laser scanners and optical sensors [J]. IAPRS, 1999, 32(3): 68-75.

[43] HAOWEN YAN, YANDONG CHU, ZHILIN LIN, et al. A quantitative description model for direction relations based on direction groups [J]. Geoinformatica, 2006,10(2): 177-196.

[44] HENGL T, WALVOORT D J J, Brown A, et al. A double continuous approach to visualization and analysis of categorical maps [J]. International Journal of Geographical Information Science, 2004, 18: 183-202.

[45] HOWARD VEREGIN, XIPING DAI. Minimizing positional error induced by line simplification [C]. Proceedings of the International Symposium on Spatial Data Quality'99:293-298.

[46] LI ZHILIN. From phenomena to essence: envisioning the nature of digital map generalization [J]. The Cartographic Journal, 1995,32 (1):45-47.

[47] LI ZHILIN. An examination of algorithms for the detection of critical points on digital cartographic lines [J]. The Cartographic Journal, 1995, 32(4):121-125.

[48] JASINSKI M J. The comparison of complexity measures for carto-

graphic lines [M]. New York: National Center for Geographic Information and Analysis, 1990:90-100.

[49] J. FLORIAN WELLMANN, KLAUS REGENAUER-LIEB. Uncertainties have a meaning: information entropy as a quality measure for 3-D geological models [J]. Tectonophysics, 2012, 526(529): 207-216.

[50] JOAO E M. Causes and consequences of map generalization [M]. London: Talyor & Francis Ltd, 1998, 10(1): 6-10.

[51] JONES C B, KIDNER D R. Database design for a multi scale spatial information system[J]. International Journal of Geographical Information Systems, 1996, 10(8):901-920.

[52] JULIE STEELE, NOAH ILIINSKY. Beautiful visualization [M]. Sebastopol, CA: O'Reilly, 2010.

[53] KILPELÄINEN T. Knowledge acquisition for generalization rules [J]. Cartography and Geographic Information Science, 2000, 27(1):41-50.

[54] KRAAK M J, ORMELING F J. Cartography visualization of spatial data [M]. Addison Wesley London, London. 2002.

[55] LAM N, QUATTROCHI D A. On the issues of scale, resolution and fractal analysis in the mapping sciences [J]. Professional Geographer, 1992, 44:88-89.

[56] LARS HARRIE. Weight-Setting and quality assessment in simultaneous [J]. Graphic Generalization, 2003, 40(3): 221-233.

[57] LIANG JIYE, LI DEYU. Information measures of roughness of knowledge and significance of attribute in rough set theory [J]. Journal of Engineering Mathematics, 2007, 17(1):106-108.

[58] LITTLE, A R. An evaluation of selected computer-assisted line simplification algorithms in the context of map accuracy standards [C]// 1989 ASPRS/ACSM Annual Conventions, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, 1989, 5: 122-32.

[59]LIU XINGJIAN, AI TINGHUA, LIU YAOLIN. Road density analysis based on skeleton partitioning for road generalization [J]. Geo-Spatial Information Science, 2009, 12(2):110-116.

[60] MACEACHREN A M, DI BIASE. Animated maps of aggregate data: conceptual and practical problems [J]. Cartography and Geographic Information Systems, 1991,18:221-229.

[61] MACEACHREN A. Time as a cartographic variable [C]// HEARNSHAW H M, UNWIN D J, CHICHESTER, WEST SUSSEX. Visualization in Geographical Information Systems. New York: Wiley and Sons. 1994:115-130.

[62]MCMASTER R B. A statistical analysis of mathematical mea-sures for linear simplification [J]. The American Carto-grapher, 1986, 13(2): 103-116.

[63]MAFFINI G M A, BITTERLICH W. Observations and comments on the generation and treatment of error in digital GIS data[J]. Geographical Analysis, 1989: 55-67.

[64]MCMASTER R B. Automated line generalization [J]. Cartographica, 1987, 27:74-111.

[65]MCMASTER R B. The geometric properties of numerical generalization [J]. Geographical Analysis, 1987, 19: 330-346.

[66]MÜLLER J C. The removal of spatial conflicts in line generalization [J]. Cartography and Geographic Information Systems, 1990,17:141-49.

[67]MÜLLER J C, WEIBEL R, LAGRANGE J P, SALGE F. Generalization: state of the art and issues [C]// MULLER J C. GIS and Generalization: Methodology and Practice. London: Taylor & Francis. 1996.

[68]VISVALINGAM M, WHYATT J D. The Douglas-Peucker algorithm for line simplification: reevaluation through visualization [C]. North-Holland . Computer Graphics Forum 9, 1990.

[69]MCCARTHY H H, HOOK J C, KNOS D S. The measurement of

association in industrial geography [D]. Iowa: State University of Iowa. 1956.

[70]MIN DENG, ZHILIN LI. A statistical model for directional relations between spatial object [J]. Geoinformatica, 2008, 12: 193-217.

[71]MOZAS A. Positional quality control in cartography using linear elements [D]. Jaén: University of Jaén. 2008.

[72]NARAYANAN R M, DESETTY M K, REICHENBACH S E. Effect of spatial resolution on information content characterization in remote sensing imagery based on classification accuracy [J]. International Journal of Remote Sensing, 2002, 23(3):537-553.

[73]NIEMANN K O, GOODENOUGH D G, HAY G J. Effect of scale on the information content in remote sensing imagery [C]. IGARSS'97, Singapore, 1997.

[74]OPENSHAW S. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modeling[J], Institure of British Geographers, Transactions, 1977, 2: 459-472.

[75]PAPADIAS D, KARACAPILIDIS N, ARKOUMANIS D. Processing fuzzy spatial queries: a configuration similarity approach[J]. International Journal of Geographical Information Science, 1999, 13(2): 93-118.

[76]PAWLAK Z. Rough set [J]. International Journal of Information and Computer Science, 1982(11): 341-365.

[77]PUPPO E, DETTORI G. Towards a formal model for multiresolution spatial maps [C]. The 4th International Symposium on Spatial and Temporal Databases SSTD 1995, Portland, M E, USA, 1995.

[78]GOYAL R K. Similarity assessment for cardinal directions between extended spatial objects [D]. Maine: The University of Maine, 2000:17-35.

[79]RANDELL D, CUI Z, COHN A. A spatial logic based on regions

and connection [C]. The 3rd International Conference on Principles of Knowledge Representation and Reasoning, San Francisco, 1992.

[80] RIPLEY B D. Modelling spatial patterns [J]. J R Stat Soc(B), 1977, 39:172-192.

[81] RIPLEY B D. Spatial statistics [M]. Chichester: Hohn Wiley, 1981.

[82] SESTER M, BRENNER M. Continuous Generalization for Visualization on Small Mobile Devices[C]. The 11th International Symposium on Spatial Data Handling, SDH 2004, Berlin, 2004.

[83] SHANNON C E, WEAVER W. The mathematical theory of communication [M]. Illinois: University of Illinois Press, 1959.

[84] SHI W Z, LIU K F. Modeling fuzzy topological relation between uncertain objects in GIS [J]. Photogrammertry Engineering and Remote Sensing, 2004, 70(8):921-930.

[85] SHI W Z, EHLERS M. "S-band", a Model to Describe Uncertainty of an Object in an Integrated GIS/Remote Sensing Environment [C]// SADA FUJIMURA, Proceedings of IGARSS' 93, Tokyo, Japan, 1993:1721-1723.

[86] SHI W Z. Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems [J]. ITC Publication, Enschede. 1994.

[87] SKIADOPOULOS S, KOUBARAKIS M. Composing cardinal direction relations [C]. Seventh International Symposium on Spatial and Temporal Databases, Los Angeles, California, USA, 2001.

[88] SKIDMORE A, TURNER B. Map accuracy assessment using line intersect sampling [J]. Photogrammetric Engineering and Remote Sensing, 1992, 58(10): 1453-1457.

[89] STELL J G, WORBOYS M F. Stratified map spaces: a formal basis for multi-resolution spatial databases[C]. The 8th Interna-

tional Symposium on Spatial Data Handling, SDH 1998, Columbia, 1998.

[90]TIMPF SABINE. Hierarchical structures in map series [D]. Vienna: Technical University of Vienna, 1998.

[91]TIMPF S, FRANCK A. A multi-scale DAG for cartographic objects [C]. Auto-Carto, Charlotte, N C,1995.

[92] TIMPF S. Hierarchical structures in map series [D]. Vienna: Technical University of Vienna,1998.

[93]TVEITE H, LANGAAS S. An accuracy assessment meted for geographical line data sets based on buffering [J]. International Journal Geographical Information Science, 1999, 14(1), 51-66.

[94]TSAI V J D. Delaunay triangulations in TIN creation: an overview and a linear-time algorithm [J]. International Journal Geographical Information Science,1993,7(6):510-524.

[95]VAN OSTEROM P. The GAP-tree, an approach to "On-the-Fly" map generalization of an area partitioning [M]. London: Taylor & Francis, 1995: 120-132.

[96]WANG S L, SHI W Z. Geographical rough space[C]. Remote Proceedings of Sensing and GIS Accuracy Assessment Symposium. Las Vegas NV, USA. 2001.

[97]WARD M, GRINSTEIN G, KEIM D. Interactive data visualization: foundations, techniques, and applications [M]. Natica Massachusetts: A K Peters Ltd. 2010.

[98]WU L X. Topological relations embodied in a generalized tri-prism (GTP) model for 3D geosciences modeling system [J]. Computers & Geosciences, 2004, 30(4): 405-418.

[99]ZADEH L A. Fuzzy sets[J]. Information and Control, 1965, 8: 338-353.

[100]ZHILIN LI, HAIGANG SUI. An integrated technique for automated generalization of contour maps [J]. The Cartographic Journal, 2000, 37(1): 29-37.

[101]ZHOU S, JONES C B. A Multi-representation spatial data model [C]. The 8th International Symposium on Spatial and Temporal Databases SSTD 2003. Santorini Island, Greece, 2003.

[102]ZHOU S, JONES C B. Design and implementation of multi-scale databases[C]. The 7th International Symposium on Spatial and Temporal Databases SSTD 2001, L A, CA, 2001.

[103]艾廷华，李精忠．尺度空间中 GIS 数据表达的生命期模型[J]．武汉大学学报：信息科学版，2010，35(7)：757-762.

[104]艾廷华，成建国．对空间数据多尺度表达有关问题的思考[J]．武汉大学学报：信息科学版，2005，30(5)：377-382.

[105]艾廷华，刘耀林．保持空间分布特征的群点化简方法[J]．测绘学报，2002，31(2)：175-181.

[106]艾廷华．基于 Monte Carlo 方法的不确定性地理现象可视化[J]．武汉大学学报：信息科学版，2004，29(3)：239-243.

[107]包磊．时空数据库中不确定性处理技术若干关键问题研究[D]．南京：南京航空航天大学，2006.

[108]蔡剑红，李德仁．多空间尺度下顾及不确定性的投影模型研究[J]．计算机工程与应用，2007，43(14)：77-80.

[109]蔡剑红，李德仁．多尺度下的不确定性空间方向锥形模型研究[J]．武汉大学学报：信息科学版，2007，32(8)：735-739.

[110]蔡永香，郭庆胜．基于 Kohonen 网络的点群综合研究[J]．武汉大学学报：信息科学版，2007，32(7)：626-629.

[111]陈军．多维动态地理空间框架数据的构建[J]．地球信息科学，2002，(1)：7-13.

[112]陈军，赵仁亮．GIS 空间关系的基本问题与研究进展[J]．测绘学报，1999，8(2)：95-100.

[113]陈棉，王秀斌，施建勇．空间信息多媒体可视化设计探讨[J]．测绘通报，2007，07：54-57.

[114]陈为，张嵩，鲁爱东，等．数据可视化的基本原理与方法[M]．北京：科学出版社，2013.

[115]陈先伟，郭仁忠，闫浩文．土地利用数据库综合中图斑拓扑关系的创建和一致性维护[J]．武汉大学学报：信息科学版，2005，30(4)：370-373.

[116]承继成，郭华东，史文中，等．遥感数据的不确定性问题[M]．北京：科学出版社，2003.

[117]崔世林，龙毅，周侗，等．一种地图点群空间分布特征的局部分维分析方法[J]．地球信息科学，2007，9(6)：25-30.

[118]邓敏，冯学智，陈晓勇．面目标间拓扑关系形式化描述的层次模型[J]．测绘学报，2005，4(2)：142-147.

[119]邓敏，冯学智，刘文宝，等．地理信息系统中拓扑关系的不确定性推理方法[J]．模式识别与人工智能，2004，17(3)：326-331.

[120]邓敏，李成名，刘文宝．利用拓扑和度量相结合的方法描述面目标间的空间关系[J]．测绘学报，2002，31(2)：164-169.

[121]邓敏，李志林，程涛．多粒度的GIS数据不确定性粗集表达[J]．测绘学报，2006，35(1)：64-70.

[122]邓敏，刘文宝，冯学智．GIS面目标间拓扑关系的形式化模型[J]．测绘学报，2005，34(1)：85-90.

[123]邓敏，马杭英．线与面目标间拓扑关系的层次表达方法[J]．测绘学报，2008，37(4)：507-520.

[124]邓敏，徐锐，李光强，等．GIS面目标间空间关系的集成表达方法[J]．计算机工程与应用，2009，45(2)：39-43.

[125]邓敏，徐锐，李志林，等．空间查询中自然语言空间关系与度量空间关系的转换方法研究：以面目标为例[J]．测绘学报，2009，38(6)：527-531.

[126]丁虹．空间相似性理论与计算模型的研究[D]．武汉：武汉大学，2004.

[127]董铁群，刘大有，王芳，等．一种基于MBR的不确定区域间方向关系建模方法[J]．电子学报，2011，39(2)：329-335.

[128] 杜道生，高文秀，龚健雅．GIS 专题数据综合的研究[J]．地理与地理信息科学，2003，19(3)：1-5.
[129] 杜世宏，王桥，魏斌，等．空间方向关系粗糙推理[J]．测绘学报，2003，32(4)：334-338.
[130] 杜世宏，王桥，杨一鹏．一种定性细节方向关系的表达模型[J]．中国图象图形学报，2004，9(12)：1496-1503.
[131] 杜世宏，王桥．不确定性空间关系[J]．中国图象图形学报，2004，9(5)：539-546.
[132] 杜晓初，熊建国．多尺度表达中空间拓扑关系等价性评价模型[J]．测绘科学，2007，32(5)：70-72.
[133] 杜晓初．多重表达中空间拓扑关系等价性研究[D]．武汉：武汉大学，2005.
[134] 范爱民，郭达志．误差熵不确定带模型[J]．测绘学报，2001，30(1)：48-53.
[135] 费立凡，何津．解决街道与建筑物图形冲突的移位模型研究[J]．武汉大学学报：信息科学版，2007，32(6)：540-543.
[136] 齐民友．概率论与数理统计[M]．北京：高等教育出版社，2002.
[137] 高建新．GIS 近期发展趋势综述[J]．测绘信息与工程，2003，28(5)：15-18.
[138] 高联雄，赵波，刘惟一．基于粗糙集的不确定区域 RCC 拓扑关系研究[J]．地理与地理信息科学，2008，24(3)：17-18.
[139] 高文秀．基于知识的 GIS 专题数据综合研究[D]．武汉：武汉大学，2002.
[140] 高振记，邬伦，杨俭．基于 RCC 及粗糙模型的模糊地理对象拓扑关系表达[J]．北京大学学报：自然科学版，2008，44(4)：597-603
[141] 龚建华，林珲，鲁学军．虚拟地理环境研究框架初探[C]．第二届“虚拟现实与地理学”学术研讨会论文集，2002. II-1-7.
[142] 郭庆胜，蔡永香．在图形简化中面状目标间拓扑关系渐进式

转换的研究[J]. 测绘科学, 2007, 32(1): 30-31.

[143]郭庆胜, 杜晓初, 刘浩. 空间拓扑关系定量描述与抽象方法研究[J]. 测绘学报, 2005, 34(2): 123-128.

[144]郭庆胜, 黄远林, 郑春燕, 等. 空间推理与渐进式地图综合[M]. 武汉: 武汉大学出版社, 2007.

[145]郭庆胜, 吕秀琴, 蔡永香, 等. 图形简化过程中空间拓扑关系抽象的规律[J]. 武汉大学学报: 信息科学版, 2008, 33(5): 520-523.

[146]郭庆胜. 地图自动综合的广义性和评价标准的研究[J]. 测绘信息与工程, 2000, 16(1): 19-21.

[147]郭仁忠, 艾廷华. 制图综合中建筑物多边形的合并与化简[J]. 武汉测绘科技大学学报, 2000, 25(1): 25-30.

[148]郭仁忠, 毋河海. 地形图上城镇居民地自动综合试验[J]. 武汉测绘科技大学学报, 1993, 18(2): 15-22.

[149]郭仁忠. 空间分析[M]. 北京: 高等教育出版社. 2001.

[150]韩志刚, 孔云峰, 秦耀辰. 地理表达研究进展[J]. 地理科学进展, 2011, 30(2): 141-148.

[151]何彬彬, 方涛, 郭达志. 空间数据挖掘不确定性及其传播[J]. 数据采集与处理, 2004, 19(4): 475-480.

[152]何建华, 刘耀林, 唐新明. 离散空间拓扑关系模型[J]. 测绘学报, 2005, 34(4): 343-348.

[153]何建华, 刘耀林, 俞艳, 等. 基于模糊贴近度分析的不确定拓扑关系表达模型[J]. 测绘学报, 2008, 37(2): 212-216.

[154]何建华, 刘耀林, 俞艳. 不确定方向关系的模糊描述模型[J]. 武汉大学学报: 信息科学版, 2008, 33(3): 257-260.

[155]何宗宜, 祝瑞国, 庞小平. 地图信息论在制图中的应用研究[J]. 地图, 1998, (2): 7-10.

[156]胡鹏, 黄杏元, 华一新. 地理信息系统教程[M]. 武汉: 武汉大学出版社, 2002.

[157]胡圣武. GIS质量评价与可靠性分析[M]. 北京: 测绘出版社, 2006.

[158]胡圣武，潘正风，王新洲，等. 地理信息系统不确定性的研究[J]. 测绘通报，2004，(9)：13-16.

[159]胡圣武，王新洲，陶本藻，等. GIS不确定性的基本理论及需解决的问题[J]. 测绘科学，2007，32(2)：15-17.

[160]胡圣武，王新洲，喻铮铮，等. 论GIS的模糊不确定性及其表示[J]. 测绘通报，2006，1：18-20.

[161]黄秋华，张新长，张文江. 土地利用数据在不同空间尺度下的精度损失分析[J]. 测绘通报，2007，(2)：44-47.

[162]黄杏元，马劲松，汤勤. 地理信息系统概论[M]. 北京：高等教育出版社，2001.

[163]姜志强. 基于GIS的地学信息的分形处理方法[J]. 吉林大学学报：信息科学版，2003，21(2)：138-142.

[164]蓝悦明，陶本藻. GIS中线元不确定性的综合量化[J]. 武汉大学学报：信息科学版，2003，28(5)：559-561.

[165]蓝悦明，陶本藻. 以点位误差描述线元位置不确定性的误差带方法[J]. 测绘学报，2004，33(4)：289-292.

[166]雷伟刚，刘大杰，童小华. GIS中曲线综合平差模型的建立[J]. 测绘学报，2005，30(3)：61-63.

[167]李大军，龚健雅，谢刚生，等. GIS中面元的误差熵模型[J]. 测绘学报，2003，32(1)：31-35.

[168]李大军，龚键雅，谢刚生，等. GIS中线元的误差熵带研究[J]. 武汉大学学报：信息科学版，2002，27(5)：462-466.

[169]李德仁，彭美云，张菊清. GIS中线要素的定位不确定性模型研究[J]. 武汉测绘科技大学学报，1995，20(4)：283-288.

[170]李德仁. 对空间数据不确定性研究的思考[J]. 测绘科学技术学报，2006，23(6)：391-392.

[171]李国重，张国芹，朱长青. GIS中基于ε_σ模型的随机折线元位置不确定性的可视化[J]. 测绘工程，2009，18(5)：10-14.

[172]李红梅. 地理空间实体类型语义相似度计算模型的研究[D].

武汉：武汉大学，2005.
[173]李精忠，艾廷华．多尺度土地利用数据库构建过程中的拓扑一致性维护[J]．测绘通报，2011，(8)：32-35.
[174]李军，庄大方．地理空间数据的适宜尺度分析[J]．地理学报，2002，57(增刊)：52-59.
[175]李霖，吴凡．空间数据多尺度表达模型及其可视化[M]．北京：科学出版社，2005.
[176]李娜，杜清运，吴小芳，等．制图综合中道路与建筑物的移位研究[J]．测绘科学，2006，31(4)：92-94.
[177]李双成，王羊，蔡运龙．复杂性科学视角下的地理学研究范式转型[J]．地理学报，2010，65(11)：1315-1324.
[178]李雯静，林志勇，龙毅．粗集分类思想在GIS点群综合中的应用[J]．武汉大学学报：信息科学版，2008，33(9)：896-899.
[179]李雯静，邱佳，龙毅．粗集方法在地图综合中的应用[J]．测绘学报，2012，41(2)：298-301.
[180]梁晔，周海燕．本体论与语义Web[J]．北京联合大学学报：自然科学版，2007，21(1)：39-44.
[181]林冬云，刘慧平．应用空间聚类进行点数据分布研究[J]．北京师范大学学报：自然科学版，2006，42(4)：419-423.
[182]刘碧龙，赵春江，吴华瑞．空间信息要素模型可视化的设计模式与应用[J]．地球信息科学，2004，04：117-120.
[183]刘春，史文中，刘大杰．GIS空间数据面元与线元不确定性的关系[J]．武汉大学学报：信息科学版，2005，30(1)：65-68.
[184]刘大杰，华慧．GIS线要素不确定性模型的进一步探讨[J]．测绘学报，1998，27(1)：45-49.
[185]刘大杰，刘春．GIS空间数据不确定性与质量控制的研究现状[J]．测绘工程，2001，10(1)：6-10.
[186]刘大杰，史文中，童小华，等．GIS空间数据的精度分析与质量控制[M]．上海：上海科学技术文献出版社，1999.

[187]刘纪芹，史开泉．基于知识含量的粗糙集不确定性度量[J]．计算机科学，2007，34(7)：171-173.

[188]刘文宝，戴洪磊，徐泮林，等．GIS中平面面位误差环的解析模型[J]．测绘学报，1998，27(4)：338-344.

[189]刘文宝，戴洪磊，徐泮林，等．平面线位误差带几何形状的解析表达[J]．测绘学报，1998，27(3)：231-237.

[190]刘文宝．空间数据不确定性理论[D]．武汉：武汉测绘科技大学，1995.

[191]刘颖，翟京生．随机点群目标空间图形的表达与识别[J]．测绘科学，2005，30(4)：39-42.

[192]鲁学军，周成虎，张洪岩，等．地理空间的尺度结构分析模式探讨，2004，23(2)：107-114.

[193]陆毅，翟京生，杜景海．数字海图点群状特征的识别、量测与综合[J]．武汉大学学报：信息科学版，2001，26(2)：134-139.

[194]罗广祥，王丹丹，王俊峰，等．线状与面状对象的拓扑关系及其量度指标体系研究[J]．测绘科学，2007，32(4)：91-92.

[195]鲁学军，周成虎，龚建华．论地理空间形象思维——空间意象的发展[J]．地理学报，1999，54(5)：401-409.

[196]吕秀琴，吴凡．多尺度空间对象拓扑相似关系的表达与计算[J]．测绘信息与工程，2006，31(2)：29-31.

[197]孟斌，王劲峰．地理数据尺度转换方法研究进展[J]．地理学报，2005，60(2)：277-288.

[198]牛继强，徐丰，姚高伟．点群多尺度表达不确定性评估方法研究[J]．测绘科学，2011，36(4)：75-77.

[199]牛继强，徐丰．空间数据多尺度表达中空间关系的统一表达模型[J]．测绘科学技术学报，2013，30(2)：182-186.

[200]彭唬，齐清文，刘兆礼，等．空间点集自动概括方法的优化设计[J]．地球信息科学，2005，7(4)：122-126.

[201]齐清文，刘岳．GIS环境下面向地理特征的制图概括的理论

和方法[J]. 地理学报，1998，53(4)：303-313.

[202]齐清文，张安定. 关于多比例尺 GIS 中数据库多重表达的几个问题的研究[J]. 地理研究，1999，18(2)：161-170.

[203]钱海忠，武芳，张琳琳，等. 基于极化变换的点群综合几何质量评估[J]. 测绘学报，2005，34(4)：361-369.

[204]钱海忠，武芳，朱鲲鹏，等. 一种基于降维技术的街区综合方法[J]. 测绘学报，2007，36(1)：102-107.

[205]施建勇，闫浩文. 点群分布要素地图综合的质量评价方法[J]. 矿山测量，2008，(5)：23-25.

[206]时晓燕. 大比例尺地形图街区式居民地分层综合模型与实现[J]. 武汉测绘科技大学学报，1993，18(增刊)：47-58.

[207]史文中，刘文宝. GIS 中线元位置不确定性的误差圆带[J]. 测绘信息与工程，1998，(2)：6-8.

[208]史文中，童小华，刘大杰. GIS 中一般曲线的不确定性模型[J]. 测绘学报，2000，29(1)：52-58.

[209]史文中，王树良. GIS 中属性不确定性的处理方法及其发展[J]. 遥感学报，2002，6(5)：393-400.

[210]史文中. 空间数据与空间分析不确定性原理[M]. 北京：科学出版社，2005.

[211]宋国民，贾奋励，王光霞. 电子地图中居民地的多尺度显示[J]. 测绘学院学报，2004，21(1)：57-60.

[212]隋春光，李霖. 本体的概念、构建及其在地理信息科学中的应用[J]. 测绘通报，2010，(4)：19-21.

[213]童小华，史文中，刘大杰. GIS 中数字化数据误差的分布检验与处理[J]. 武汉测绘科技大学学报，2000，25(1)：79-84

[214]王家耀，成毅. 空间数据的多尺度特征与自动综合[J]. 海洋测绘，2004，24(4)：1-3.

[215]王家耀. 空间信息系统原理[M]. 北京：测绘出版社，2001.

[216]王建华. 基于多媒体技术的空间信息可视化研究[D]. 武汉：武汉大学，1999.

[217] 王杰臣，张伟，毛海城，等. GIS 网络分析的图简化方法研究[J]. 测绘学报，2001，30(3)：263-268.

[218] 王劲峰. 空间分析[M]. 北京：科学出版社，2006.

[219] 王明常，应申，李霖. 基于 Voronoi 图的空间信息多尺度表达[J]. 吉林大学学报：地球科学版，2005，35(4)：539-542.

[220] 王明孝，张志华，杨维芳. 地理空间数据可视化[M]. 北京：科学出版社，2012.

[221] 王桥，毋河海. 地图图斑群自动综合的分形方法研究[J]. 武汉测绘科技大学学报，1996，21(1)：59-63.

[222] 王桥，吴纪桃. 制图综合方根规律模型的分形扩展[J]. 测绘学报，1996，25(2)：104-109.

[223] 王少一，王昭，杜清运. 顾及地图要素级别的几何信息量量测方法[J]. 测绘科学，2007，32(4)：60-62.

[224] 王树良，李德仁，史文中，等. 地学粗空间的理论与应用[J]. 武汉大学学报：信息科学版，2002，27(3)：274-282.

[225] 王新洲，汤仲安，陈志辉. GIS 中不规则样条拟合曲线的 ε_m 模型带[J]. 武汉大学学报：信息科学版，2004，29(1)：58-62.

[226] 王艳慧，陈军，蒋捷. GIS 中地理要素多尺度概念模型的初步研究[J]. 中国矿业大学学报，2003，32(4)：376-382.

[227] 王艳慧，孟浩. GIS 中地理要素多尺度表达间层次连通性的研究[J]. 湖南科技大学学报：自然科学版，2006，21(1)：59-63.

[228] 王延亮，刘登杰. 离散目标群的边界搜索[J]. 测绘科学，2007，32(5)：101-102.

[229] 王瑜，胡运发，张凯. 基于粗集理论的知识含量度量研究[J]. 计算机研究与发展，2004，41(9)：1500-1506.

[230] 王昭，费立凡. 基于语法层的地图综合信息量原则[J]. 测绘科学，2007，32(6)：21-23.

[231] 魏保峰. GIS 空间数据中线元不确定性及可视化研究[D]. 昆

明：昆明理工大学，2008.
[232]魏海平. GIS中多尺度地理数据库的研究与应用[J]. 测绘学院学报，2000，17(2)：134-137.
[233]邬伦，高振记，史文中，等. 地理信息系统中的不确定性问题[M]. 北京：电子工业出版社，2010.
[234]邬伦，于海龙，高振纪，等. GIS不确定性框架体系与数据不确定性研究方法[J]. 地理学与国土研究，2002，18(4)：1-5.
[235]毋河海. 分维扩展的数值试验研究[J]. 武汉测绘科技大学学报，1998，23(4)：329-336.
[236]毋河海. 凸壳原理在点群综合中的应用[J]. 测绘工程，1997，6(1)：1-6.
[237]毋河海. 自动综合的结构化实现[J]. 武汉测绘科技大学学报，1996，21(3)：277-285.
[238]毋河海. GIS环境下城市平面图形的自动综合问题[J]. 武汉测绘科技大学学报，2000，25(3)：196-202.
[239]吴凡. 基于小波分析的线状特征数据无级表达[J]. 武汉大学学报：信息科学版，2004，29(6)：488-491.
[240]吴纪桃，王桥. 小波分析在GIS线状数据图形简化中的应用研究[J]. 测绘学报，2000，29(1)：71-75.
[241]谢俊平，杨敏华. 拓扑关系和方向关系的统一表达模型[J]. 测绘科学，2012，37(2)：146-148.
[242]谢琦，刘大有，虞强源，等. 一种不确定区域间的方向关系模型[J]. 吉林大学学报：理学版，2006，44(5)：748-753.
[243]徐丰，牛继强. 空间数据多尺度表达不确定性研究进展[J]. 信阳师范学院学报：自然科学版，2012，25(2)：276-280.
[244]徐丰，张兴国，牛继强. 线状要素多尺度表达质量评价模型研究[J]. 测绘科学，2009，34(1)：95-97.
[245]徐丰，牛继强. 线状要素多尺度表达不确定性研究[J]. 信阳师范学院学报：自然科学版，2007，20(4)：525-528.
[246]徐丰，牛继强，姚高伟. 基于位置不确定性的多尺度空间方

向关系相似性粗集模型[J]. 信阳师范学院学报：自然科学版，2013，26(3)：366-369.

[247]徐庆荣，杜道生．计算机地图制图原理[M]. 武汉：武汉测绘科技大学出版社，1993.

[248]闫浩文，褚衍东．多尺度地图空间相似关系基本问题研究[J]. 地理与地理信息科学，2009，25(4)：42-44.

[249]尹连旺，李京．GIS 中基本要素的无级比例尺数据处理技术研究[J]. 北京大学学报：自然科学版，1999，35(6)：842-849.

[250]张景雄，杜道生．位置不确定性与属性不确定性的场模型[J]. 测绘学报，1999，28(3)：244-249.

[251]张景雄．空间信息的尺度、不确定性与融合[M]. 武汉：武汉大学出版社，2008.

[252]郑春燕，胡华科．多尺度地图表达居民地与道路间拓扑关系一致性的模糊评价[J]. 测绘科学，2012，1：1-5.

[253]周平，唐新明，张过．多时相统计数据空间动态可视化模型研究[J]. 武汉大学学报：信息科学版，2012，09：1130-1133.

[254]朱长青，张国芹，史文中．2 维线元不确定性 ε_{σ} 模型误差带几何特征的代数研究[J]. 测绘学报，2007，36(4)：463-467.

[255]祝国瑞．地图学[M]. 武汉：武汉大学出版社，2004.

[General Information]
书名=空间数据多尺度表达的不确定性分析模型
作者=徐丰，牛继强著
页数=265
SS号=13522494
DX号=
出版日期=2014.04
出版社=武汉大学出版社